

Outer Billiards on the Penrose Kite: Compactification and Renormalization

Richard Evan Schwartz *

January 12, 2013

Abstract

We give a fairly complete analysis of outer billiards on the Penrose kite. Our analysis reveals that this 2 dimensional dynamical system has a 3-dimensional compactification, a certain polyhedron exchange map defined on the 3-torus, and that this 3-dimensional system admits a renormalization scheme. The two features, the compactification and the renormalization scheme, allow us to make sharp statements concerning the distribution, large- and fine-scale geometry, and hidden algebraic symmetry, of the orbits. One concrete result is that the union of the unbounded orbits has Hausdorff dimension 1. We establish many of the results with computer-aided proofs that involve only integer arithmetic.

1 Introduction

1.1 Background

Outer billiards is a dynamical system defined relative to a convex shape in the plane. B.H. Neumann [N] introduced outer billiards in the late 1950s, and J. Moser [M1] popularized the system as a toy model for celestial mechanics. See [T1], [T2], and [DT1] for expositions of outer billiards and many references.

* Supported by N.S.F. Research Grant DMS-0072607

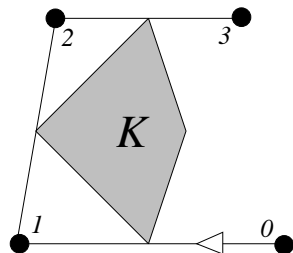


Figure 1.1: outer billiards relative to K .

To define an outer billiards system, one starts with a bounded convex set $K \subset \mathbf{R}^2$ and considers a point $x_0 \in \mathbf{R}^2 - K$. One defines x_1 to be the point such that the segment $\overline{x_0x_1}$ is tangent to K at its midpoint and K lies to the right of the ray $\overrightarrow{x_0x_1}$. The iteration $x_0 \rightarrow x_1 \rightarrow x_2 \dots$ is called the *forwards outer billiards orbit* of x_0 . It is defined for almost every point of $\mathbf{R}^2 - K$. The backwards orbit is defined similarly.

Outer billiards is an affinely natural system, in the sense that an affine map $T : P \rightarrow Q$ carrying the convex shape P to the convex shape Q also carries the outer billiards orbits relative to P to the outer billiards orbit relative to Q . The reason is that affine maps carry line segments to line segments and respect the property of bisection.

One of the central questions about outer billiards is the *Moser-Neumann question*, which asks if an outer billiards system can have unbounded orbits. Here is an abbreviated list of work on this problem.

- J. Moser [M2] sketches a proof, inspired by KAM theory, that outer billiards on K has all bounded orbits provided that ∂K is at least C^6 smooth and positively curved. R. Douady [D] gives a complete proof.
- In Vivaldi-Shaïdenko [VS], Kolodziej [Ko], and Gutkin-Simanyi [GS], it is proved (each with different methods) that outer billiards on a *quasirational polygon* has all orbits bounded. This class of polygons includes polygons with rational vertices and regular polygons. In the rational case, all orbits are periodic.
- In [T2], Tabachnikov shows the existence of aperiodic orbits in the regular pentagon case, and works out a renormalization scheme to explain their structure.
- D. Genin [G] shows that all orbits are bounded for the outer billiards systems associated to trapezoids. See §1.

- In [S1], we settled Moser-Neumann question by showing that outer billiards has some unbounded orbits when defined relative to $K(\phi^{-3})$. Here ϕ is the golden ratio and $K(A)$ denotes the kite with vertices

$$(-1, 0); \quad (0, 1) \quad (0, -1) \quad (A, 0); \quad (1)$$

Figure 1.1 shows an example.

- In [S2], we showed that outer billiards has unbounded orbits relative to $K(A)$, when A is any irrational number in $(0, 1)$.
- Dolgopyat and Fayad [DF] showed that outer billiards has unbounded orbits relative to the half-disk and other “caps” made from slicing a disk nearly in half.

The shapes in [DF] and [S2] are the only ones known to produce unbounded orbits, though certainly it now seems that unboundedness is a common phenomenon. Our monograph [S2] gives more details about the history of the problem.

The set $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$ is invariant under the dynamics on $K(A)$. We call the orbits that lie in this set *special*. In [S2] we gave quite a lot of information about special orbits on kites. For instance, we gave a formula for the Hausdorff dimension of the set $U_1(A)$ of unbounded special orbits, in terms of something akin to the continued fraction expansion of A . As a special case,

$$\dim(U_1(\phi^{-3})) = \frac{\log(2)}{\log(\phi^3)}. \quad (2)$$

We also showed that every unbounded special orbit is self-accumulating. This is to say that every point of an unbounded special orbit O is an accumulation point of O .

In the 300 page [S2] we only considered the special orbits, for the sake of “brevity”. There is quite a bit more to say about the general orbits, and our purpose here is to say some of it, at least for the Penrose kite. The first phenomenon is that outer billiards on the Penrose kite, an unbounded 2 dimensional system, has a 3 dimensional compactification. We saw similar things in [S1] and [S2].

The second phenomenon is that this higher dimensional compactification has a renormalization scheme. The renormalization scheme is the main new feature of this paper. Its existence allows us to get some precise results about the dynamics. We think that a similar scheme exists in great generality, but we currently don’t have any techniques for investigating it in general.

1.2 The Distribution of Unbounded Orbits

In all that follows, it goes without saying that our results concern only outer billiards on the Penrose kite.

Theorem 1.1 *Every orbit is either periodic or unbounded in both directions, and the union of unbounded orbits has Hausdorff dimension 1.*

It is convenient to state our remaining results in terms of the square of the outer billiards map, which we call ψ . The map ψ leaves invariant the set

$$\mathbf{R}_y^2 = \bigcup_{n \in \mathbf{Z}} \mathbf{R} \times (y + 2n). \quad (3)$$

Each \mathbf{R}_y^2 is a discrete countable family of horizontal lines. We always take y as a point of the circle $\mathbf{R}/2\mathbf{Z}$. In [S1] and [S2] we studied the orbits in \mathbf{R}_1^2 .

Let $\mathbf{Z}[\phi]$ denote the ring of elements $m + n\phi$ where $m, n \in \mathbf{Z}$. We will sometimes use the notation

$$\begin{bmatrix} m \\ n \end{bmatrix} = m + n\phi. \quad (4)$$

We define an equivalence relation on points of $\mathbf{R}/2\mathbf{Z}$. We say that $a \sim b$ if

$$b = \pm\phi^{3k}a + 2m + 2n\phi; \quad k, m, n \in \mathbf{Z}. \quad (5)$$

We will explain this equivalence relation in a more natural way in §1.4

An even length increasing sequence $a_1 < \dots < a_{2n}$ canonically defines a Cantor set C , as follows. Let T_k be the similarity carrying $[a_1, a_{2n}]$ to $[a_{2k-1}, a_{2k}]$. Then C is the limit set of the semigroup generated by T_1, \dots, T_n . For instance, the sequence $0 < 1/3 < 2/3 < 1$ defines the usual middle-third Cantor set. We let $C^\#$ denote the set obtained from C by removing the endpoints of all the complementary regions.

Theorem 1.2 *The set \mathbf{R}_y^2 contains unbounded orbits if and only if $y \sim c$ and $c \in C^\#$, where C is the Cantor set defined by the sequence*

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} < \begin{bmatrix} 2 \\ -1 \end{bmatrix} < \begin{bmatrix} 4 \\ -2 \end{bmatrix} < \begin{bmatrix} 6 \\ -3 \end{bmatrix} < \begin{bmatrix} -2 \\ 2 \end{bmatrix} < \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The set of such y is a dense set of Hausdorff dimension $\log(3)/\log(\phi^3)$.

1.3 The Distribution of Periodic Orbits

We define the *winding number* of a periodic orbit p to be half the number of times $\{\psi^n(p)\}$ intersects the strip

$$\Sigma = \mathbf{R} \times [-2, 2] \quad (6)$$

This definition makes sense geometrically. It turns out that, at least far from the origin, the ψ -orbits generally wind around the origin, nearly following a large octagon, and return to Σ after each half-revolution. See §4.2 and §4.2. The following result says that orbits of high winding number are extremely pervasive. In particular, it says that \mathbf{R}_y^2 has periodic orbits of arbitrarily high winding number provided that $y \notin 2\mathbf{Z}[\phi]$.

Theorem 1.3 *For any integer N , there is a finite subset $W_N \subset 2\mathbf{Z}[\phi]$ such that \mathbf{R}_y^2 has periodic orbits of winding number greater than N if $y \notin W_N$.*

As with any polygonal outer billiards system, every periodic point is contained in a maximal convex polygon consisting of points which all have the same period and combinatorial behavior. We call these maximal polygons *periodic tiles*. For convenience, we include K itself as a periodic tile. We call the union of the periodic tiles the *dynamical tiling*, even though it is only a tiling of a subset of the plane, and we denote it \mathcal{D} . Theorem 1.1 says that \mathcal{D} fills up everything but 1-dimensional set.

A clean, finitary description of \mathcal{D} is a beyond our reach. \mathcal{D} is quite complicated. In particular, it contains infinitely many different shapes. However, we will get a near-complete understanding of the portion of \mathcal{D} contained in the triangle T with vertices

$$(0, 1); \quad (0, \phi^{-3}); \quad (\phi^{-1}, \phi). \quad (7)$$

T is bounded by lines extending 3 of the sides of K , as in Figure 1.2. We call T the *fundamental triangle*.

Figure 1.2 below shows a tiling \mathcal{T} of a full measure subset of T by an infinite union of kites and octagons. the kites are all similar to each other and the octagons are all similar to each other. The similarity factors all have the form ϕ^{3k} , where ϕ is the golden ratio and k is an integer. We will describe \mathcal{T} precisely in §3.1. The fractal set which is the complement of the polygonal tiles has Hausdorff dimension 1. The set of points in this fractal set, having well defined orbits, also has Hausdorff dimension 1.

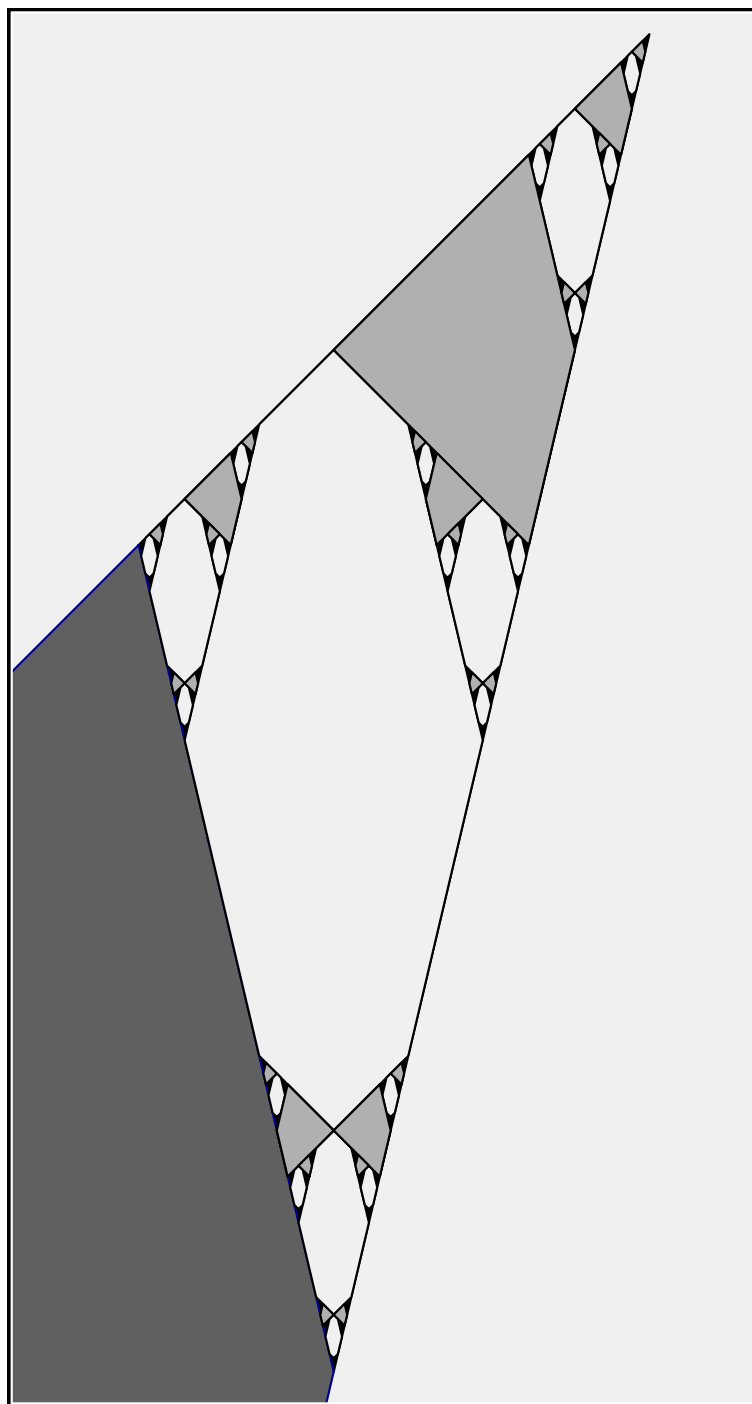


Figure 1.2: The triangle T and its tiling \mathcal{T} .

Theorem 1.4 *Suppose that $p \in T$ has a well defined orbit. Then p has a periodic orbit if and only if p lies in the interior of a tile of \mathcal{T} . Each tile of \mathcal{T} is partitioned into finitely many periodic tiles. For any N , only finitely many of the tiles contain periodic orbits having winding number less than N .*

Remarks:

- (i) Theorem 1.4 is not quite as sharp as we would like. Experimental evidence suggests that each tile of \mathcal{T} is itself a periodic tile. Indeed, we plotted Figure 1.2 using our outer billiards program, which finds the periodic tiles.
- (ii) The set y such that the horizontal line of height y intersects $T - \mathcal{T}$ in an infinite set is precisely the Cantor set from Theorem 1.2. This is where the Cantor set comes from.
- (iii) The fractal set of aperiodic points for outer billiards on the regular pentagon has a self-similar structure akin to the one in Figure 1.2. In that case, the aperiodic orbits are all bounded. See Tabachnikov [T2] for details.

1.4 Renormalization

Now we reconsider the equivalence relation defined in connection with Theorem 1.2. To say that $a \sim b$ means that $b = T(a)$, where T is an affine map defined over $\mathbf{Z}[\phi]$ whose multiplying coefficient is a unit in $\mathbf{Z}[\phi]$ and whose action is congruent to the identity mod $2\mathbf{Z}[\phi]$. Such maps form a group, and act on $\mathbf{R}/2\mathbf{Z}$ with dense orbits. The equivalence classes we defined in connection with Theorem 1.2 are precisely the orbits of this group action.

We think of G_2 as related to Γ_2 , the *level 2 congruence subgroup* of $PSL_2(\mathbf{Z})$, the modular group. In [S2] we discovered that, with respect to the special orbits on arbitrary kites, there is a kind of hidden Γ_2 -symmetry. For instance, the dimension of the set of unbounded orbits on \mathbf{R}_1^2 , as a function of the kite parameter, is a Γ_2 -invariant function. We were not able to see the kind of renormalization structure that we establish here, but the way we think of things is that the union of all the dynamical systems defined by outer billiards on kites is a kind of plane-bundle over the parameter interval. We think there is a group $\widehat{\Gamma}_2$ acting (in way that meaningfully relates to the dynamics) on this bundle in such a way that Γ_2 gives the action on the base space and our group G_2 here is the restriction of $\widehat{\Gamma}_2$ to a particular fiber.

Having indulged in some speculation, we now return to concrete results.

Our renormalization works best for orbits which we call *generic*. Say that a point $p = (x, y) \in \mathbf{R}^2$ is *generic* if it does not satisfy any equation of the

form

$$ax + by + c = 0; \quad a, b, c \in \mathbf{Z}[\phi]; \quad a \neq 0. \quad (8)$$

That is, (x, y) is generic if it does not lie on a non-horizontal line that is defined over $\mathbf{Z}[\phi]$. The outer billiards map is entirely defined on the set of generic points, and preserves this set. So, it makes sense to speak of a generic orbit. Theorem 1.9 below explains the sense in which we do not miss much by ignoring the non-generic orbits. In §5.8 we discuss the issues surrounding the renormalization of non-generic orbits.

We define two kinds of equivalence relations between orbits.

- Let $\langle O \rangle$ denote the graph of an orbit $O = \{p_n\}$, namely the subset $\{(n, p_n)\} \subset \mathbf{R}^3$. We call two orbits O_1 and O_2 *coarsely equivalent* if there is a K bi-lipschitz map $h : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ such that $h(\langle O_1 \rangle)$ and $\langle O_2 \rangle$ lie in K -tubular neighborhoods of each other. We call K the *coarse equivalence constant*.
- We call the orbits O_1 and O_2 *locally equivalent* if, for each point $p_1 \in O_1$, there is a point $p_2 \in O_2$, open disks Δ_1 and Δ_2 , and a similarity $S : \Delta_1 \rightarrow \Delta_2$ such that $S(p_1) = S(p_2)$ and S conjugates the first return map $\psi|_{\Delta_1}$ to one of the two return maps $\psi^{\pm 1}|_{\Delta_2}$, at least on generic points. In particular, $S(O_1 \cap \Delta_1) = O_2 \cap \Delta_2$. Also, S carries $\mathcal{D} \cap \Delta_1$ to $\mathcal{D} \cap \Delta_2$ modulo the operation of subdividing each tile into finitely many smaller polygons.

Theorem 1.5 *Suppose that $y_1 \sim y_2$ are two parameters in the same G_2 orbit. Then there is a bijection between a certain subset of the orbits in $\mathbf{R}_{y_1}^2$ and a certain subset of the orbits in $\mathbf{R}_{y_2}^2$. These subsets contain all generic unbounded orbits and also all generic periodic orbits having sufficiently high winding number. Corresponding orbits are both locally and coarsely equivalent, and the coarse equivalence constant only depends on (y_1, y_2) and not on the individual orbit.*

The last statement in Theorem 1.5 is important for the periodic orbits. Every two periodic orbits are coarsely equivalent, so we need some kind of uniformity in order to make a meaningful statement.

Here are two applications of Theorem 1.5.

Theorem 1.6 *Every generic unbounded orbit is self-accumulating in at least one direction.*

Theorem 1.6 is probably true for the non-generic unbounded orbits as well, but our techniques fall a bit short of this result.

Theorem 1.7 *Any generic unbounded orbit is locally and coarsely equivalent to a generic unbounded orbit that intersects the fundamental triangle T . In particular, in a small neighborhood of any generic point p with an unbounded orbit, the dynamical tiling in a neighborhood p is isometric to a neighborhood of \mathcal{T} , modulo the addition or removal of a countable set of lines.*

The ambiguity concerning the countable set of lines comes from the set containing the non-generic orbit. The main thing that is missing in Theorem 1.7 is a description of the dynamical tiling in the neighborhood of points that do not have well-defined orbits. Figure 1.3 shows what the dynamical tiling looks like in a certain region whose lowest vertex is $(3, 0)$, a point which turns out to be a fixed point of renormalization in a sense that we will make precise in our Fixed Point Theorem from §5.

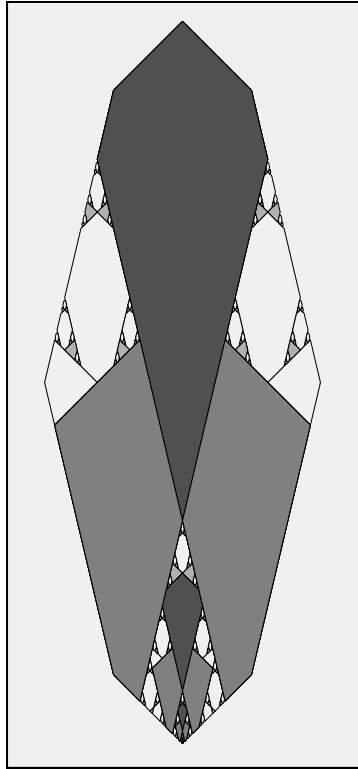


Figure 1.3: The dynamical tiling near $(3, 0)$.

1.5 Fine Points

In this section we present a few more result that seem a bit more specialized than the ones above.

For all $r > 0$, let Y_r denote those $y \in Y$ such that \mathbf{R}_y^2 has an unbounded orbit that intersects the disk of radius r about the kite vertex $(\phi^{-3}, 0)$.

Theorem 1.8 *For all $r > 0$ the set Y_r is a nowhere dense set having Hausdorff dimension $\log(3)/\log(\phi^3)$.*

Comparing Theorem 1.8 and Theorem 1.2, we can say qualitatively that most unbounded orbits stay far from the origin. In particular, for any compact subset, we can find an unbounded orbit that avoids this set. This is much different than what happens for the special orbits. For instance, our result [S2, Erratic Orbit Theorem] says in particular that every unbounded orbit in \mathbf{R}_1^2 returns infinitely often to every neighborhood of the kite vertex $(0, 1)$.

Our next result quantifies the sense in which there are fewer non-generic unbounded orbits than there are generic unbounded orbits.

Theorem 1.9 *The union of non-generic unbounded orbits has Hausdorff dimension at most $\log(3)/\log(\phi^3)$, and \mathbf{R}_y^2 contains non-generic unbounded orbits only if it contains generic unbounded orbits.*

Here is one more result about periodic orbits.

Theorem 1.10 *Suppose that $y = m + n\phi$ with m and n odd integers. Then there is some open neighborhood V of the line $\mathbf{R} \times \{y\}$ that contains no unbounded orbits.*

Theorem 1.10 says that these special lines are contained in the interior of the closure of the union of the periodic tiles. One might say that these lines are paved over with periodic tiles.

Here is a nice special case of Theorem 1.2.

Theorem 1.11 *Suppose that $y = m + n\phi$ with m and n integers. Then \mathbf{R}_y^2 has unbounded orbits if and only if m is odd and n is even.*

We wonder which $y \in \mathbf{Q}[\phi]$ are such that \mathbf{R}_y^2 has unbounded orbits. Theorem 1.2 reduces this to an arithmetical question about the Cantor set C .

1.6 Discussion

Outer billiards is a non-compact 2-dimensional system. We exhibit a 3 dimensional compactification of (a certain first return map of) the outer billiards system. The compactification turns out to be a polyhedron exchange map defined on a 3-torus $\widehat{\Sigma}$. This result is very similar to the Master Picture Theorem in [S2]. We call our result the Compactification Theorem.

Once we find the compactification, we will observe that it has a renormalizable structure that is responsible for most of the results presented above. Specifically, we will find sets $\widehat{A}, \widehat{B} \subset \widehat{\Sigma}$ together with a 3 to 1 map $\widehat{R} : \widehat{A} \rightarrow \widehat{B}$ such that \widehat{R} conjugates the first return map on \widehat{A} to the first return map on \widehat{B} . Our main result here is Theorem 5.3, the Renormalization Theorem.

It is worth comparing the renormalization scheme here to the one [T2] for regular pentagons. In that case, one picks a certain bounded subset B of the plane and observes that the first return map to B is renormalizable. What this means, in part, is that there is a subset $A \subset B$ and a similarity $R : A \rightarrow B$ which conjugates f_A to f_B . Here f_A is the first return map to A and f_B is the first return map to B . It also means that the subsets A and B are large enough to capture all the dynamical properties of the whole map.

A direct renormalization like this would be very difficult to establish in the presence of unbounded orbits, because the first return times to any sufficiently large compact set would be unbounded. Also, thanks to Theorem 1.8, there really is no compact set that “captures” all the relevant dynamics. What we do for the penrose kite is show the existence of the renormalization scheme on a higher dimensional compactification. Once we make the compactification, the renormalization is at least vaguely similar to what happens for the regular pentagon.

One difference between the renormalization scheme here and the one for the regular pentagon is that the one here involves a 3-to-1 covering map rather than a similarity. Another difference is that the regular pentagon case can be analyzed by hand, just following the orbits of several convex polygons. Here we need to keep track of about a million polyhedra just to see that the scheme works. This is what seems to make a computer-aided proof necessary.

The moral of the story is that if one wants to find renormalization schemes for polygonal outer billiards, one should first compactify. In some sense, this is a lesson I learned from John Smillie. When hearing about my earlier work on outer billiards, Smillie guessed that probably there was a renormalization scheme behind it.

1.7 Overview of the Paper

In §2 we study a certain 3-to-1 covering map $R : \mathbf{R}/2\mathbf{Z} \rightarrow \mathbf{R}/2\mathbf{Z}$ which governs the structure of our renormalization map \widehat{R} . This map is closely related to the Cantor set in Theorem 1.2. The work in §2 is a microcosm for the rest of the paper: We give some theoretical arguments to reduce the main result to a finite calculation which is too big to do by hand, and then we explain a rigorous computer calculation that finishes the proof.

In §3 we study the tiling \mathcal{T} shown in Figure 1.2 and prove a number of results about it.

In §4 we present some background information about outer billiards and polyhedron exchange maps. Most of this information also appears in [S1] and [S2].

In §5 we state the Compactification Theorem, the Renormalization Theorem, as well as several useful auxiliary results. One of these auxiliary results, the Fundamental Orbit Theorem, explains the structure of Figure 1.2. Another result, the Fixed Point Theorem, explains the dilation symmetry in Figure 1.3. The remaining results, the Near Reduction Theorem and the Far Reduction Theorem, explain the sense in which renormalization “brings orbits closer to the origin”.

In §6 we put together the material from §2-5 to deduce all the theorems mentioned in the introduction. We warn the reader that the order we prove these results is rather different from the way we have stated them. For instance, it takes almost every other result in order to prove Theorem 1.1.

In §7 we explain the main computational ideas we use in our proofs. These algorithms perform fairly standard tasks – e.g. detecting whether a point is contained in the interior of a polyhedron. Later chapters will refer back to the methods explained in §6. Given that the computational algorithms perform fairly standard tasks, the reader won’t lose much understanding of the overall proof if they just skim the material in §7. On the other hand, we think that §7 might be very useful for someone who would like to reproduce our results or to prove similar results in a related setting.

In §8 we prove a technical result, known as the Pinwheel Lemma, which is helpful in proving the Compactification Theorem. Versions of the Pinwheel Lemma also appear in [S1], [S2], and [S3], and related results appear in other works on the subject by other authors.

In §9 we prove the Compactification Theorem. Our proof is very similar to the proof of [S1, Arithmetic Graph Lemma].

In §10 we prove the Remormalization Theorem by reducing it to an explicit calculation involving the polyhedron exchange map.

In §11 we prove the Fundamental Orbit Theorem and the Fixed Point Theorem.

In §12 we prove the Near Reduction Theorem.

In §13 we prove the Far Reduction Theorem.

In §14 we include coordinates for all the polyhedra involved in the polyhedron exchange map, as well as coordinates the sets \hat{A} and \hat{B} .

1.8 Computational Issues

The general strategy of our paper is to reduce all the results to statements about finite partitions of various sets by polygons and polyhedra. These partitions sometimes involve a huge number of individual pieces, on the order of a million, and so it is necessary for us to use the computer to deal with them effectively.

One source of potential error in a computer aided proof is floating-point (or roundoff) error. To avoid any possibility of floating-point errors, we perform our calculations using exact arithmetic in the number ring $\mathbf{Z}[\phi]$. The special nature of our constructions allows us to do this. With exact arithmetic calculations, the one potential hazard is overflow error. We avoid overflow error by checking the sizes of the integers involved after every arithmetic operation. We describe the main features of these calculations in §7.1.

We illustrate our structural results in detail in §4, but there is probably no way for the reader to appreciate the details of the objects involved in without seeing explicit (and interactive) computer plots. We made an extremely detailed and extensive java applet that lets the reader see everything in the paper. This applet is available on my website. For the reader who would like to do his/her own experiments, we include enough information in the appendix so that in principle one could reproduce the calculations.

We would like to comment on some of the figures in the paper. To illustrate certain definitions which make sense for any kite $K(A)$, we will draw $K(1/4)$ in place of $K(\phi^{-3})$, because $K(1/4)$ is much easier to draw by hand. We draw these pictures mainly to give the reader a picture of what is going on, and for these purposes a picture of $K(1/4)$ tells the whole story. Note that $1/4 - \phi^{-3} = .0139\dots$, so the pictures we draw for $K(1/4)$ are geometrically quite close to the ones for $K(\phi^{-3})$. On the other hand, the computer pictures we draw will show $K(\phi^{-3})$.

1.9 Further Results

There are some other things I've noticed about outer billiards on the Penrose kite. The other things have a different character from the results here. They have to do with the patterns one sees in the so-called arithmetic graph associated to the dynamics. In §4.5 I briefly discuss one of the results, something I call the *freezing phenomenon*. I would like to have presented some of these results, this paper is already long enough and also I have not worked out proofs for these other results.

1.10 Acknowledgements

I would especially like to thank John Smillie, who has for several years repeatedly told me to “get renormalization into the picture” of outer billiards. I would also like to thank Gordon Hughes, Rick Kenyon, Curt McMullen, and Sergei Tabachnikov for helpful conversations about matters related to this work.

2 The Circle Renormalization Map

In this chapter, we define and then study the renormalization map R that we mentioned in the introduction. We also define and study the tiling \mathcal{T} we mentioned in the introduction and plotted in Figure 1.2. The map R and the tiling \mathcal{R} are closely related.

2.1 Basic Definition

Let

$$\mathbf{T} = \mathbf{R}/2\mathbf{Z} \tag{9}$$

Here we give a precise definition of the renormalization map $R : \mathbf{T} \rightarrow \mathbf{T}$ discussed in the introduction.

We decompose $[0, 2]$ into 5 intervals.

- $I_1 = [0, \phi^{-2}] = [0, 2 - \phi]$.
- $I_2 = [\phi^{-2}, 2\phi^{-2}] = [2 - \phi, 1 - \phi^{-3}]$.
- $I_3 = [2\phi^{-2}, 2 - 2\phi^{-2}] = [1 - \phi^{-3}, 1 + \phi^{-3}]$.
- $I_4 = [2 - 2\phi^{-2}, 2 - \phi^{-2}] = [1 + \phi^{-3}, \phi]$.
- $I_5 = [2 - \phi^{-2}, 2] = [\phi, 2]$.

We define R as follows.

- If $y \in I_1$ then $R(y) = \phi^3 y$. Note that $R(I_1) = I_1 \cup \dots \cup I_4$.
- If $y \in I_2$ then $R(y) = y + \phi - \phi^{-2}$. Note that $R(I_2) = I_5$.
- If $y \in I_3$ then $R(y) = \phi^3 y - \phi^3 + 1$. Note that $R(I_3) = I_1 \cup \dots \cup I_5$.
- If $y \in I_4$ then $R(y) = y - \phi + \phi^{-2}$. Note that $R(I_4) = I_1$.
- If $t \in I_5$ then $R(y) = \phi^3 y - 2\phi^3 + 2$. Note that $R(I_5) = I_2 \cup \dots \cup I_5$.

The map R pieces together correctly on the endpoints of these intervals, and induces a degree 3 covering map $R : \mathbf{T} \rightarrow \mathbf{T}$.

Note that R preserves the ring $\mathbf{Z}[\phi]$, because $\phi^k \in \mathbf{Z}[\phi]$ for all $k \in \mathbf{Z}$. Note also that the second iterate R^2 is strictly expanding.

2.2 The First Descent Lemma

As we mentioned above, R preserves the set

$$\mathbf{A} = \mathbf{Z}[\phi] \cap [0, 2] \quad (10)$$

The goal of this section is to prove the following number-theoretic result.

Lemma 2.1 (Descent I) *The action of R on \mathbf{A} is the identity mod $2\mathbf{Z}[\phi]$, and for any $y \in \mathbf{A}$, there is some n such that $R^n(y)$ is one of:*

$$0; \quad 1; \quad \phi, \quad 2 - \phi; \quad -1 + \phi, \quad -3 + 3\phi, \quad 3 - \phi, \quad 5 - 3\phi.$$

Each collection of points, grouped according to the parity of their coefficients, forms a periodic cycle for R .

One can see directly from the formulas that R is the identity mod $2\mathbf{Z}[\phi]$. We concentrate on the proof of the second statement.

Given $y = m + n\phi \in \mathbf{A}$ we define

$$N(y) = \max(|m|, |n|). \quad (11)$$

Lemma 2.2 *If $y \in \mathbf{A} \cap (I_2 \cup I_4)$ then $N(R(y)) \leq N(y) + 2$.*

Proof: If $y = m + n\phi \in I_2$ then $R(y) = (m - 2) + (n + 2)\phi$. In this case the result is obvious. If $y \in I_4$, the proof is very similar. ♠

Lemma 2.3 *If $y \in \mathbf{A} \cap (I_1 \cup I_3 \cup I_5)$ then $N(R(y)) < N(y)/2 + 8$.*

Proof: Suppose first that $y \in I_1$. Let $y = m + n\phi$. We compute

$$R(y) = (1 + 2\phi)(m + n\phi) = (m + 2n) + (2m + 3n)\phi.$$

Now we observe the following two inequalities.

$$|m + 2n| = |(m + \phi n) - (\phi - 2)n| \leq 2 + \phi^{-2}|n| < 2 + N(y)/2$$

$$|2m + 3n| = |(2m + 2\phi n) - (2\phi - 3)n| \leq 4 + \phi^{-3}|n| < 4 + N(y)/2.$$

The last inequalities come from the fact that $y \in [0, 2]$. When y lies in I_3 or I_5 the argument is the same, except that we have must either add 2 or 4 to our estimate to account for the translational part of R . ♠

Corollary 2.4 *For any $y \in [0, 2]$ there is some n such that $N(R^n(y)) < 20$.*

Proof: Suppose that $y \in \mathbf{A}$ has the property that $N(y) > 20$. Note that $R(y)$ and $R^2(y)$ cannot both lie in $I_2 \cup I_4$. Hence, we may combine our last two results to establish the bound $N(R^2(y)) < y/2 + 10 < N(y)$. So $R^2(y)$ has smaller integer norm than does y . ♠

The above work reduces the proof of the Descent Lemma I to a finite calculation of what happens to those y for which $N(y) < 20$. We make the finite calculation and see that the Descent Lemma I holds for these values as well. Indeed, in the next section, we will describe a much more extensive calculation.

We mention a variant which we will need in one place. Let S be the involution $S(y) = 2 - y$.

Lemma 2.5 *Let $y \in \mathbf{Z}[\phi]$ and let $y' = R_1 \circ \dots \circ R_N(y)$, where each R_k is either R or $R \circ S$. Then y' is one of the values listed in the Descent Lemma I provided that N is sufficiently large.*

Proof: The map R commutes with S and all the cycles listed in the Descent Lemma I are invariant under S . (The first cycle is really $[0] = [2]$.) The result follows immediately from these two facts. ♠

2.3 The Second Descent Lemma

Let G_2 be the affine group defined in §1.4. This group has two components, so to speak. We let G_2^+ denote the index 2 subgroup consisting of maps of the form

$$T(x) = \phi^{3k}x + b; \quad k \in \mathbf{Z}; \quad b \in 2\mathbf{Z}[\phi]. \quad (12)$$

This formula differs from the one in Equation 5 only in that we do not allow a minus sign in front of ϕ^{3k} .

Lemma 2.6 (Descent II) *Two elements $y_1, y_2 \in \mathbf{R}/2\mathbf{Z}$ lie in the same G_2^+ orbit iff there are positive integers n_1 and n_2 such that $R^{n_1}(y_1) = R^{n_2}(y_2)$.*

Proof: All the maps defining R lie in G_2^+ . So, if there are positive integers n_1 and n_2 such that $R^{n_1}(y_1) = R^{n_2}(y_2)$ then y_1 and y_2 lie in the same G_2^+ orbit. The converse is the interesting direction.

For the converse, suppose that y_1 and y_2 lie in the same G_2^+ orbit. Considering the action of R , we can replace y_2 by some image $y'_2 = R^k(y_2)$ such that $y_1 - y'_2 \in 2\mathbf{Z}[\phi]$. So, without loss of generality, we can consider the case where we already know that $y_1 - y_2 \in 2\mathbf{Z}[\phi]$. The Descent Lemma I proves this result whenever our points lie in $\mathbf{Z}[\phi]$. So, it suffices to consider the case when neither point belongs to $2\mathbf{Z}[\phi]$.

We find it more convenient to work with a new map that is closely related to R . We define $\rho = R$ on $I_1 \cup I_3 \cup I_5$ and $\rho = R^2$ on $I_2 \cup I_4$. Unlike R , which pieces together continuously across the endpoints of the intervals in the partition, the map ρ is not defined on the endpoints. However, the points we are considering, and their orbits, never hit these endpoints. It suffices to prove this result for ρ in place of R .

To describe the map ρ , we let $f[m, n]$ denote the map

$$x \rightarrow \phi^3 x + m + n\phi. \quad (13)$$

Then

- On I_1 , we have $\rho = f[0, 0]$.
- On I_2 , we have $\rho = f[2, -2]$.
- On I_3 , we have $\rho = f[0, -2]$.
- On I_4 , we have $\rho = f[-2, -2]$.
- On I_5 , we have $\rho = f[0, -4]$.

This is a short calculation which (after many tries) we did correctly. We omit the details.

We write $\langle y_1, y_2 \rangle = \max(|m|, |n|)$, where $y_1 - y_2 = m + n\phi$. The same argument as in Lemma 2.3 and Corollary 2.4 shows (with tons of room to spare) that

$$\langle y_1, y_2 \rangle > 100 \quad \implies \quad \langle y'_1, y'_2 \rangle < \langle y_1, y_2 \rangle - 1. \quad (14)$$

Here we have set $y'_k = \rho(y_k)$. Equation 14 reduces this lemma to the case when $\langle y_1, y_2 \rangle \leq 100$. In the next section, we will explain our computer-assisted proof that The Descent Lemma holds for such choices. ♠

2.4 A Dynamical Computation

Notice that our statement that the Descent Lemma II holds for pairs (y_1, y_2) with $\langle y_1, y_2 \rangle \leq 100$ is not *obviously* a finite calculation, because it involves infinitely many values. However, we will explain how to reduce the problem to a finite calculation, which we then make. This situation is typical of the results in this paper. The challenge is reduce seemingly infinite statements to finite computations.

Let $Q = [0, 2]^2$. The set of pairs (y_1, y_2) of interest to us lie on a finite number of line segments of slope 1 that are contained in Q . By switching the order of the two points if necessary, we can assume that $y_1 < y_2$. We let ρ act on Q by having ρ act separately on each coordinate.

The square Q is partitioned into 25 subsquares

$$Q_{ij} = I_i \times I_j \tag{15}$$

on which ρ is entirely defined and a similarity. Our point of view is that ρ acts separately on each Q_{ij} , and the action on the various boundaries depends on which square we include the boundary in. Anyway, we don't care about what happens on the boundaries: As we said above, the points we consider, and their orbits, never hit the boundaries.

Say that a *diagonal* is a segment of slope 1 contained in Q . We call the diagonal *small* if it is contained in one of the 25 subsquares. The endpoints of a small diagonal might lie in the boundary of the subsquare, but this is fine with us. Given a small diagonal I , we can define $\rho(I)$ using the action of $\rho|_{Q_{ij}}$. Note that $\rho(I)$ is another diagonal, but not necessarily a small one.

Each diagonal has a canonical decomposition into small diagonals: We just take the intersections with the 25 sub-squares. This we have a kind of dynamical system defined on lists of diagonals: Given a list of diagonals, we first subdivide each member of the list into small diagonals. Next, we let ρ act on all the small diagonals. And so on. As one final nicety, we switch the two coordinates of the endpoints, if necessary, so that all our diagonals lie about the line $y_2 - y_1 = 0$. (We do this simply for computational convenience.)

Say that a *good seed* is a diagonal of the form

$$\Delta(m, n) = Q \cap \{y_2 - y_1 = 2m + 2n\phi\}, \tag{16}$$

where

$$2m + 2n\phi \in [0, 2]; \quad \max(|2m|, |2n|) \leq 100. \tag{17}$$

We run the dynamical system starting with any good seed and we find that, after finitely many steps, the only remaining intervals lie in

$$\Delta(0, 0) \cup \Delta(-2, 2) \cup \Delta(4, -2). \quad (18)$$

Points $(y_1, y_2) \in \Delta(0, 0)$ obviously satisfy $y_1 = y_2$. When $(y_1, y_2) \in \Delta(-2, 2)$, it means that $y_2 - y_1 = \phi - \phi^{-2}$. In this case, we either have $y_1 \in I_2$ and $y_2 = R(y_1)$ or $y_2 \in I_4$ and $y_1 = R(y_2)$. Compare the definition of the map R .

It only remains to deal with those points in $\Delta(4, -2)$. For this purpose, we just have analyze the dynamics more carefully. Define

$$\Delta'(4, -2) = \Delta(4, -2) \cap Q_{24}. \quad (19)$$

When we perform the dynamics on $\Delta(4, -2)$ we find that the following occurs.

- $\Delta(4, -2)$ breaks up into 5 small diagonals, one of which is $\Delta'(4, -2)$.
- ρ maps each of the 4 other small diagonals into $\Delta(0, 0) \cup \Delta(-2, 2)$.
- ρ maps $\Delta'(4, -2)$ back into $\Delta(4, -2)$.

This analysis shows that the Renormalization Lemma can only fail for a pair of points (y_1, y_2) such that $\rho^n(y_1) \in I_2$ for all n (and also $\rho^n(y_2) \in I_4$ for all n .) But the fixed point of $\rho_n|_{I_2}$ is an endpoint of I_2 and ρ is an expanding map. Since y_1 is not this endpoint, we see that $\rho^n(y_1)$ eventually escapes I_2 , and we are done.

Remarks:

- (i) We perform the calculations with exact arithmetic, as explained in §7.1.
- (ii) In §7.2 we explain how we eliminate any possibility of overflow error in our calculations. Even without specific guards against overflow error (which we do have) for all our seeds the dynamical system reaches the 3 end-states above very quickly and all integers remain pretty small.

2.5 The Cantor Set

In this section, we give some information about the Cantor set C from Theorem 1.2. First of all, the main property of C is that both C and $C^\#$ are forward R -invariant. That is, $R(C) = C$ and $R(C^\#) = C^\#$. Indeed, this is how we discovered the map R .

Lemma 2.7 *Let $y \in \mathbf{Z}[\phi]$. Then $y \in C^\#$ only if $y \equiv 1 \pmod{2\mathbf{Z}[\phi]}$ and $y \in C - C^\#$ only if $y = m + n\phi$ with m even.*

Proof: Of the 8 values listed in the conclusion of the Descent Lemma I, we see that only 1 lies in $C^\#$. Given any $y \in \mathbf{Z}[\phi]$ which intersects $C^\#$, we simply note that $R^n(y) \in C^\#$ for all n . But there is some n such that $R^n(y)$ is one of the 8 values listed in the Descent Lemma I. But this means that $R^n(y) = 1$. Hence $y \sim 1 \pmod{G_2^+}$. But this means that $y \equiv 1 \pmod{2\mathbf{Z}[\phi]}$.

For the second statement, we just have to rule out the case that $y \equiv -1 + \phi \pmod{2\mathbf{Z}[\phi]}$. But we check easily that $-1 + \phi \notin C$. But C is forward R -invariant and $R^n(y) = -1 + \phi$ for some n . This situation is impossible. ♠

To each point $y \in \mathbf{R}/2\mathbf{Z}$ we assign a *renormalization sequence* in the digits $\{1, 2, 3, 4, 5\}$. The sequence is such that the n th term is k if and only if $R^n(y) \in I_k$. Not every point has a unique renormalization sequence. A point y has a non-unique renormalization sequence if and only if $R^n(y) \in \partial I_k$ for some n and some k . Examining the endpoints of our intervals and also the conclusion of the Descent Lemma, we see that this happens if and only if $y = m + n\phi$, where m and n are integers and n is even. In particular, all points of $C^\#$ have unique renormalization sequences.

Lemma 2.8 *A point lies in $C^\#$ if and only if its renormalization sequence is unique and has no 5's in it.*

Proof: A point y has a non-unique renormalization sequence if and only if $R^n(y) \in \partial I_k$ for some n and some k . Examining the endpoints of our intervals and also the conclusion of the Descent Lemma, we see that this happens if and only if $y = m + n\phi$, where m and n are integers and n is even. In particular, all points of $C^\#$ have unique renormalization sequences.

Note that the renormalization sequence of a point in $y \in C^\#$ cannot start with 5, because $I_5 = [\phi, 2]$ only shares its bottom endpoint with C , and this endpoint lies in $C - C^\#$. Since $C^\#$ is forward invariant, we see that the renormalization sequence cannot have any 5's in it at all.

Now we know that $C^\#$ only contains points that have unique renormalization sequences with no 5's in them. Conversely, suppose y has a unique renormalization without 5's. Suppose that $y \notin C$. Since $y \notin I_5$, we can say that y lies on one of the bounded components of $\mathbf{R} - C$. These components

all have diameter ϕ^{-3k+1} for $k = 1, 2, 3 \dots$. The largest component has size ϕ^{-2} , and is precisely the interior of I_2 . But $y \notin I_2$ because then $R(y) \in I_5$. But, iteration of R increases the sizes of all gaps except the largest one. Hence $R^n(y)$ lies in the largest gap for some n . This is a contradiction.

The endpoints of the gaps do not have unique renormalization sequences. So, the same argument rules out the possibility that $y \in C - C^\#$. We conclude that $y \in C^\#$. ♠

3 The Fundamental Tiling

3.1 Definition of the Tiling

Let \mathcal{T} be the tiling of the fundamental triangle T shown in Figure 1.2. In this section we define \mathcal{T} precisely. As we mentioned in the introduction, T is bounded by three sides of the Penrose kite K . The reader can most easily understand the definitions we make by referring back to Figure 1.2.

Let $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be the map that fixes the top vertex of T and shrinks distances by a factor of ϕ^3 . We define $K_0 = f(K)$, where K is the Penrose kite. K_0 is the largest kite in the tiling \mathcal{T} . The largest octagon J_0 in \mathcal{T} has 4-fold dihedral symmetry, and 3 vertices located at

$$(13 - 8\phi, 4 - 2\phi) \quad (5 - 3\phi, 6 - 3\phi) \quad (-3 + 2\phi, -2 + 2\phi). \quad (20)$$

This is enough information to characterize J_0 uniquely. A calculation shows that

$$T - K_0 - J_0 = (T_{11} \cup T_{12}) \cup T_{31} \cup T_{32} \cup T_{41} \quad (21)$$

Here T_{ij} is a similar copy of T , with scaling factor ϕ^{-3} . Our notation is such that $T_{jk} \subset I_j$, the interval used in the definition of the map R . Just to pin thing down exactly, we say that the centroid of T_{i1} lies to the left of the centroid of T_{i2} . The bottom triangles T_{11} and T_{12} , mirror images of each other, are not disjoint. However, the similarities carrying T to T_{11} and T_{12} both map J_0 to the same smaller octagon. Therefore, we can compatibly subdivide each of our smaller triangles into a kite, an octagon, and a union of 5 smaller triangles. Continuing this process forever, we get the tiling \mathcal{T} .

Let S denote the union of points in T that do not belong to the interiors of any of the tiles of \mathcal{T} . Let $S^\#$ denote those points of S that do not belong to the boundary of any tile of \mathcal{T} . Then $S = S^\#$ is contained in a countable union of lines defined over $\mathbf{Z}[\phi]$. These are the lines extending the sides of the boundaries of the tiles in \mathcal{T} . We call $S^\#$ the *fundamental fractal*. We have the basic relation

$$C^\# = \pi_2(S^\#); \quad C = \pi_2(S). \quad (22)$$

Here π_2 is projection onto the second coordinate, and C is the Cantor set from Theorem 1.2.

3.2 Hausdorff Dimension

The purpose of this section is to prove the following result. Here $\dim(S)$ refers to the Hausdorff dimension of S . See [F] for details about Hausdorff dimension.

Lemma 3.1 (Dimension) *Let S be the fundamental fractal and let L be any countable collection of non-horizontal lines. Then $\dim(S - L) = 1$. In particular, $\dim(S^\#) = 1$.*

Proof: After we remove J_0 and K_0 from T we are left with the union in Equation 21. We have the equation

$$T_{11} \cup T_{12} - R_{12}^{-1}(K_0) = T_{11} \cup T'_{12}, \quad (23)$$

where T'_{12} is similar to T , with similarity factor ϕ^{-6} . Figure 3.1 shows this operation.

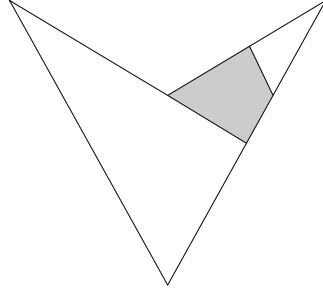


Figure 3.1: $(T_{11} \cup T_{12}) - R_{12}^{-1}(K_0) = T_{11} \cup T'_{12}$.

So, removing two open kites and an open octagon from T leaves the 5 (temporarily renamed) disjoint triangles T_1, \dots, T_5 . We have

$$\sum_{k=1}^5 \text{diam}(T_k) = \phi^{-3}(4 + \phi^{-3})\text{diam}(T) = \text{diam}(T) \quad (24)$$

The set S is the self-similar fractal which is the limit set of the semigroup generated by the similarities which carry T to each of the smaller triangles. From here it is an exercise to show that $\dim(S) = 1$.

Note that C , the Cantor set from Theorem 1.2 is the set $y \in [0, 2]$ such that the line of height y intersects S . Each line in L intersects S in a set of dimension $\dim(C) < 1$. Since $\dim(S) = 1$ and $\dim(S \cap L) < 1$, we have $\dim(S - L) = 1$. ♠

3.3 The Horizontal Intersections

Let $C^\#$ be as in the previous chapter. Every $y \in C^\#$ is such that

$$L_y = \mathbf{R} \times \{y\} \quad (25)$$

intersects $S^\#$ nontrivially. The goal of this chapter is to prove the following result.

Lemma 3.2 (Horizontal) *For any $y \in C^\#$, the set $L_y \cap S$ is a Cantor set. Hence $L_y \cap S^\#$ is obtained from a Cantor set by removing at most countably many points.*

We begin by explaining how the set S may be constructed recursively, as the nested intersection of finite unions of (overlapping) triangles.

Recall that $T - J_0 - K_0 = \bigcup T_{ij}$, a union of 5 triangles. Let ρ_{ij} be the similarity such that $\rho_{ij}(T_{ij}) = T$. Define $S_0 = T$ and (inductively) S_n such that

$$\rho_{ij}(S_n \cap T_{ij}) = S_{n-1}. \quad (26)$$

Then S_n is a finite union of triangles and

$$S = \bigcap S_n. \quad (27)$$

Our proof involves an analysis of how these triangles sit with respect to the horizontal lines.

Remarks:

(i) Note the similarity between the maps ρ_{ij} and the map ρ considered in the previous chapter. precisely, ρ_{ij} acts on the horizontals having heights in I_i exactly as ρ acts on I_i . We will pursue this analogy further in the next section.

(ii) We have to be careful with our definition, because T_{11} and T_{12} overlap. Here is the justification for what we do. Assume by induction that S_{n-1} is well-defined and that $S_{n-1} \cap (T_{11} \cap T_{12})$ has bilateral symmetry and that ρ_{11} and ρ_{12} have the same action on this intersection. Then S_n is well-defined and inherits all these same properties by symmetry.

Let $\{k_n\}$ be the renormalization sequence associated to y . By Lemma 2.8 this sequence has no 5's. Since S_n is a finite union of triangles, the intersection

$$\Lambda_n = L_y \cap S_n \quad (28)$$

is a finite union of disjoint intervals.

Lemma 3.3 *Let J be a maximal interval of Λ_n . Then $J \cap \Lambda_{n+1}$ is nonempty and contains 2 maximal intervals in case $k_n = 3$ or $(k_n, k_{n+1}) = (1, 4)$.*

Proof: Let $y_0 = y$ and $y_n = R^n(y)$ and $L(n) = L_{y_n}$. Our sequence $\{k_n\}$ starts with k_0 . We have $y_n \in I_{k_n}$.

Recall that $S_0 = T$. It follows from induction and fact that $\rho_{ij}(T_{ij}) = T$ that there is a sequence of maps $\rho_0, \dots, \rho_{n-1}$ such that

$$\rho_{n-1} \circ \dots \circ \rho_0(J \cap \Lambda_{n+m}) = L(n) \cap S_m; \quad m = 0, 1, \dots \quad (29)$$

Here $\rho_j = \rho_{k_j, m_j}$, where $m_j \in \{1, 2\}$. The sequence $\{\rho_i\}$ is not necessarily unique, because T_{11} and T_{12} overlap. We don't mind this.

Since $k_n \in \{1, 3, 4\}$, the set on the right hand side of Equation 29 is a nontrivial union of intervals. Hence $J \cap \Lambda_{n+1}$ has at least one interval. When $k_n = 3$ the set $L(n) \cap S_1$ contains 2 intervals. Hence $J \cap \Lambda_{n+1}$ contains 2 intervals in this case. When $k_n = 1$ and $k_{n+1} = 4$ the line $L(n)$ lies above the top vertex of $T_{11} \cap T_{12}$. Hence $L(n) \cap S_1$ again consists of 2 intervals. Hence, so does $J \cap \Lambda_{n+1}$. ♠

Corollary 3.4 *Let J be a maximal interval of Λ_n . Then $J \cap \Lambda_m$ contains at least 2 disjoint intervals for m sufficiently large.*

Proof: The sequence associated to y cannot terminate in an infinite string of 1's. Otherwise, there is another sequence associated to y which terminates in an infinite string of 5's. So, the associated sequence either has an infinite number of 3's or an infinite number of $(1, 4)$'s. The corollary now follows immediately from Lemma 3.3. ♠

It follows from Corollary 3.4 that the nested intersection $\bigcap \Lambda_n$ is a Cantor set. This completes the proof of the Horizontal Lemma.

3.4 Notation

We have already mentioned the similarities ρ_{ij} . These maps have the property that $\rho_{ij}(T_{ij}) = T$. These maps would seem to suit us perfectly well, but it turns out that there is a slightly more elaborate collection of maps that are better adapted to the structure of outer billiards. We introduce these maps here. These maps have the advantage that they are all homotheties.

Define the reflection in the vertical line $x = 1$:

$$\eta(x, y) = (2 - x, y). \quad (30)$$

We first change our notation a bit. We let $T^+ = T$ and $T_{ij}^+ = T$, etc. That is, we attach the (+) superscript to all the objects associated to the fundamental triangle. Next, we define $X^- = \eta(X^+)$ for any object X .

The two triangles T^+ and T^- are mirror images of each other. We will see eventually that the dynamical tiling intersects T^- exactly in the tiling \mathcal{T}^- . It turns out that the dynamics on $T^+ \cup T^-$ works out more nicely than the dynamics on just T^+ , even though ultimately all our results are phrased just in terms of T^+ .

We define the following ten maps.

- $R_{11}^+ = \eta_{11} : T_{11}^+ \rightarrow T^+$.
- $R_{12}^+ = \eta \circ \rho_{11} : T_{11}^+ \rightarrow T^-$.
- $R_{31}^+ = \eta \circ \rho_{31} : T_{31}^+ \rightarrow T^-$.
- $R_{32}^+ = \eta_{32} : T_{32}^+ \rightarrow T^+$.
- R_{41}^+ is the isometry carrying T_{41}^+ to T_{12}^+ .
- $R_{ij}^- = \eta \circ R_{ij}^+ \circ \eta$.

The maps R_{ij}^- acts similarly to the map R_{ij}^+ . For instance $R_{31}^-(T_{31}^-) = T^+$.

Recall that the renormalization map R equals the map R_i on the interval I_i . The maps R_{ij}^\pm are all similarities preserve the horizontal foliation and act on the horizontal lines intersecting T_{ij}^\pm as R_i acts on I_i . In particular, the maps R_{41}^\pm are better adapted to R_4 than the map ρ_{41} , which is really adapted to the map $R_1 \circ R_4$.

The maps R_{41}^\pm are orientation preserving isometries. The remaining 8 maps are orientation reversing similarities, with expansion constant ϕ^3 .

3.5 The Renormalization Set

Given subsets $A, B \subset T^+ \cup T^-$, we write $A \rightarrow B$ if $A \subset T_{ij}^\pm$ and $B = R_{ij}^\pm(A)$. We use this definition in particular for points. Suppose that (p_1, q_1) is a pair of points, both at the same height. We write $(p_1, q_1) \rightarrow (p_2, q_2)$ if $p_1 \rightarrow p_2$ and $q_1 \rightarrow q_2$. We write $p \sim q$ if

$$(p, q) \rightsquigarrow \dots \rightsquigarrow (p', q'); \quad p' = q'. \quad (31)$$

Define

$$\Upsilon(p) = \{q \mid q \sim p\}. \quad (32)$$

Note that $\Upsilon(p)$ consists of points that are all on the same horizontal level as p . We call $\Upsilon(p)$ the *renormalization set* of p .

Remark: In §6.6, we will see that p and q lie in the same outer billiards orbit provided that p and q both have unbounded orbits and $p \sim q$. This is a step in our proof that generic unbounded orbits are self-accumulating.

Let $C^\#$ be the set studied in the previous chapter. The goal of this section is to prove the following result.

Lemma 3.5 (Density) *Let $\Lambda^\pm = L \cap S^\pm$, where L is a horizontal line whose height lies in $C^\#$. For any point $p \in \Lambda^+ \cup \Lambda^-$, the set $\Upsilon(p)$ is dense in $\Lambda^+ \cup \Lambda^-$.*

the Density Lemma. We call T^+ and T^- the *distinguished triangles* of depth 0. Recall that S^\pm is contained in the nested intersection of sets S_n^\pm . We say that a *distinguished triangle* is a maximal triangle of S_n^\pm . There are 2×5^n distinguished triangles. We call n the *depth* of the distinguished triangle.

We say that a distinguished triangle is *related* to a renormalization set if it intersects the horizontal line containing the renormalization set. Let $P(n)$ be the property that every renormalization set intersects each related distinguished triangle depth n . It suffices to prove that the statement $P(n)$ is true for all n .

If τ is any distinguished triangle of depth n , then $\tau \rightarrow \tau' \rightarrow \tau''$ where one of τ' or τ'' has depth $n - 1$. From this, and from the definitions, we see that $P(n - 1)$ implies $P(n)$. It just remains to establish $P(0)$.

Say that points $p, q \in T^+ \cup T^-$ are *distantly placed* if one of the points lies in T^+ and the other lies in T^- . Say that $y \in C^\#$ is *good* if $P(0)$ holds for all renormalization sets of height y .

Lemma 3.6 *Suppose that $y_1 \in R^{-1}(y_2)$ and y_2 is good. Then y_1 is good.*

Proof: Let p_1 be some point having height y_1 . We have $p_1 \rightarrow p_2$ for some p_2 having height y_2 . Since y_2 is good, there is some q_2 such that $p_2 \sim q_2$ and p_2, q_2 are distantly placed. Without loss of generality, assume that $p_2 \in T^+$. There is a depth 1 distinguished triangle τ^+ such that $\tau^+ \rightarrow T^+$ and $p_1 \in \tau^+$. Let $\tau^- = \rho(\tau^+)$. Then $\tau^- \rightarrow T^-$ and points in τ^- are distantly placed from points in τ^+ . In particular, we can find $q_1 \in \tau^-$ such that $q_1 \rightarrow q_2$. But then p_1 and q_1 are distantly placed and $p_1 \sim q_1$. Since p_1 was chosen arbitrarily, y_1 is good. ♠

Lemma 3.7 *Suppose that $y \in C^\# \cap I_3$. Then y is good.*

Proof: Any horizontal line having a height in I_3 intersects the 4 disjoint triangles T_{3j}^\pm for $j = 1, 2$. Let ρ be the union of the 4 special maps. Then

$$\rho(T_{31}^+) = \rho(T_{32}^-) = T^+; \quad \rho(T_{32}^+) = \rho(T_{31}^-) = T^-.$$

It follows from this equation that every renormalization set of height y either intersects both of (T_{31}^+, T_{32}^-) or both of (T_{32}^+, T_{31}^-) . ♠

Lemma 3.8 *Suppose that $y \in C^\# \cap I_1 \cap R^{-1}(I_4)$. Then y is good.*

Proof: Any horizontal line having a height in I_3 intersects the 4 triangles T_{1j}^\pm for $j = 1, 2$, in disjoint intervals. (This is true even though the triangles themselves are not disjoint.) The rest of the proof is as in Lemma 3.7. ♠

Let $y \in C^\#$ be arbitrary. As in the proof of Corollary 3.4, there must be some n such that $R^n(y)$ satisfies either Lemma 3.7 or Lemma 3.8. But then $R^n(y)$ is good. But then y is good as well. Hence, $P(0)$ holds. But then $P(n)$ holds for all n . This completes the proof of the Density Lemma.

4 Preliminaries

4.1 Polytope Exchange Maps

Definition: For us, a *polytope exchange map* is a quadruple (M, X_1, X_2, Ψ) , where M is a flat manifold (possibly with boundary), X_1 and X_2 are locally finite partitions of M into convex polytopes, and $f : M \rightarrow M$ is a piecewise isometric bijection which carries X_1 to X_2 . We mean that f is a translation when restricted to each polytope P of X_1 and $f(P)$ is a polytope of X_2 . Technically, f is not defined on the boundaries of the polytopes. When M is a compact manifold, we require that the partitions be finite.

As a special case, suppose that $M = \mathbf{R}^n / \Lambda$, where $\Lambda \subset \mathbf{Z}^n$ is a discrete group of translations. We say that a polytope in M is *golden* if any lift to \mathbf{R}^n has all vertices with coordinates in the ring $\mathbf{Z}[\phi]$. This definition is independent of lift. We call an associated polytope exchange map *golden* if all the polytopes are golden, and if all the translations are defined by vectors in $(\mathbf{Z}[\phi])^n$. In this paper we will consider two golden polytope exchange maps.

- A 2 dimensional non-compact polygon exchange Ψ , whose domain is the infinite strip $\Sigma = \mathbf{R} \times [-2, 2]$.
- A 3 dimensional compact polyhedron exchange $\hat{\Psi}$, whose domain is the torus $\hat{\Sigma} = (\mathbf{R}/2\mathbf{Z})^3$.

Fibered Polyhedron Exchange Maps: Let \mathcal{H} be the foliation of $\hat{\Sigma}$ by horizontal 2-tori – those obtained by holding the third coordinate constant. We say that a leaf of \mathcal{H} is *golden* if its height lies in $\mathbf{Z}[\phi]$. We call a golden polyhedron exchange map on $\hat{\Sigma}$ *fibered* if it preserves the leaves of \mathcal{H} and if the restriction to each golden leaf is a golden polygon exchange map. The second condition does not follow automatically from the first; there is an auxilliary condition we need on the edges of the polyhedron partition, as we now explain.

Let e be a non-horizontal edge of some polyhedron in the partition. Let (x_1, y_1, z_1) and (x_2, y_2, z_2) be the endpoints of e . The 4 quantities

$$\frac{x_2 - x_1}{z_2 - z_1}, \quad \frac{y_2 - y_1}{z_2 - z_1}, \quad \frac{x_2 z_1 - x_1 z_2}{z_2 - z_1}, \quad \frac{y_2 z_1 - y_1 z_2}{z_2 - z_1} \quad (33)$$

all belong on $\mathbf{Z}[\phi]$ if and only if every nontrivial intersection of e and a golden leaf has coordinates in $\mathbf{Z}[\phi]$. We omit the easy proof of this result.

Periodic Tiles: Let (M, X_1, X_2, f) be a polytope exchange map, not necessarily a golden one. Given a periodic point $p \in M$, there is a maximal convex polytope D_p consisting of points q such that p and q have the same period and the same itinerary. By *itinerary* we mean the sequence $\{n_k\}$, defined so that the (n_k) th polytope $P(n_k)$ of X_1 contains $f^k(p)$. We have

$$D_p = \bigcap_k f^{-k}(P(n_k)) \quad (34)$$

Being the finite intersection of convex polytopes, D_p is also a convex polytope. We call D_p a *periodic tile*. We call the union of all the periodic tiles the *dynamical tiling*. When all points of M are periodic, the dynamical tiling is another partition of M .

Our terminology is slightly misleading. In general, the dynamical tiling only fills a subset of M . This subset need not even be dense in M . A nice example is furnished by outer billiards on a generic trapezoid. See [G], where the issue is phrased somewhat differently. In our examples, thanks to the results in this paper, the dynamical tilings are dense.

Homogeneity: As one might expect, the dynamical tiling has some nice homogeneity properties. We say that two points $x_1, x_2 \in M$ are *tile isometric* if there are open disks Δ_1 and Δ_2 such that

- $x_j \in \Delta_j$ for $j = 1, 2$.
- $\mathcal{D} \cap \Delta_1$ is isometric to $\mathcal{D} \cap \Delta_2$.

Note that x_j need not be the center of Δ_j and the isometry from $\mathcal{D} \cap \Delta_1$ to $\mathcal{D} \cap \Delta_2$ need not carry x_1 to x_2 . We call a subset $S \subset M$ *tile homogeneous* if every pair of points in M are tile isometric.

As a related definition, we say that $x_1, x_2 \in M$ are *tile similar* if the above properties hold, with the word *similar* replacing the word *isometric*.

Lemma 4.1 *Every orbit is tile homogeneous.*

Proof: Let p_0 and $p_n = f^n(p_0)$ be two points in the orbit. Let $g = f^n$. The map g is an isometry in a neighborhood of p_0 and g conjugates f to itself. It follows from this fact that p_0 and p_n are tile isometric. ♠

4.2 The Return Map

The square outer billiards map ψ is a local translation. For points far from the origin, the dynamics of ψ is quite simple. forward iterates of ψ generally circulate counterclockwise around the kite K . Near the origin, the dynamics of ψ is slightly more complicated. Our paper [S3] considers these dynamics carefully, for fairly general convex polygons.

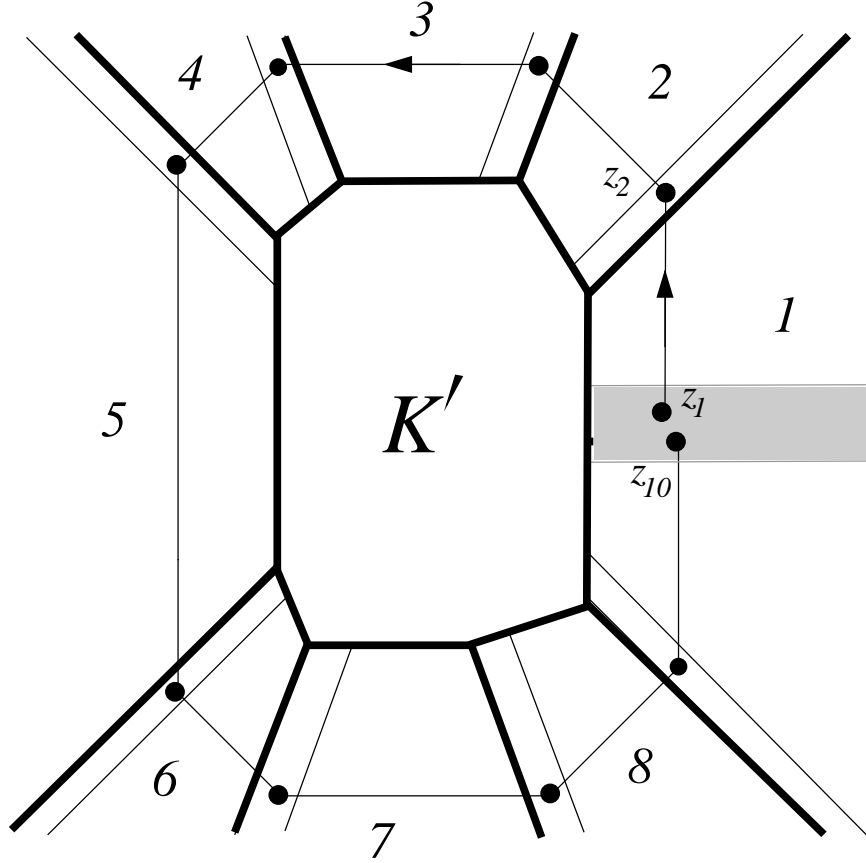


Figure 4.1: The second return map far from the origin

The set K' shown in Figure 4.1 is a large compact set that contains K well in its interior. Given that the ψ -orbits generally circulate around the kite K , at least far from K , it makes sense to consider the return map to each half of a suitable horizontal strip. Define

$$\Sigma = \mathbf{R} \times [-2, 2]. \quad (35)$$

half this strip is shaded in Figure 4.1. What makes this strip canonical is that, far from K and near the x -axis, consecutive iterates of ψ differ by the vector $(0, \pm 4)$. So, Σ has just the right width.

Remark: Sometimes we will want to leave off the bottom boundary of Σ and sometimes we won't. It turns out that the first return map is the identity on the boundary of Σ , so we can essentially just ignore these points.

The definition of the return map to “each half” of Σ presents some problems for points near K . There are “bad points” of Σ that are too close to K . These bad points are in the regions labelled B in Figure 4.2.

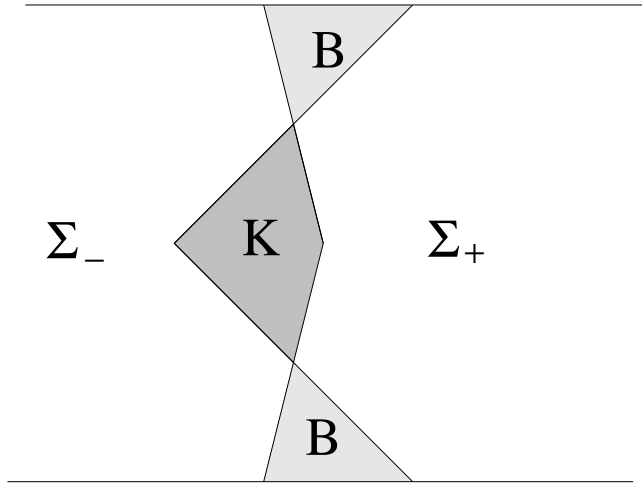


Figure 4.2: A decomposition of Σ into regions.

The problem with the bad points is that one cannot really say which side of K they are on. We mention this problem in order to justify a modified version of the return map, which we now define. Let Σ_- and Σ_+ be the two components of $\Sigma - K - B$. Let $\Psi_{\pm} : \Sigma_{\pm} \rightarrow \Sigma_{\pm}$ to be the first return map of ψ to Σ_{\pm} . Just to be clear, Ψ_+ is the first return to Σ_+ and Ψ_- is the first return to Σ_- . We define $\Psi = \Psi_+ \cup \Psi_-$ to be the “union map”, defined on $\Sigma_+ \cup \Sigma_-$. Finally, to get a map on all of Σ , we define Ψ to be the identity on $K \cup B$. Note that this definition does not correspond with the action of outer billiards on B , but we will deal specially with the points in B whenever necessary.

We call $\Psi : \Sigma \rightarrow \Sigma$ the *return map*.

Lemma 4.2 (Return) *The following is true.*

1. *Every ψ orbit intersects $\Sigma_+ \cup \Sigma_-$.*
2. *Ψ is well defined on all points of Σ that have a well-defined orbit.*
3. *There is some C such that $|\Psi(p) - p| < C$ for all $p \in \Sigma$ with a well defined orbit.*

Proof: First of all, we proved the same result in [S2, §2.3], in the context of special orbits (on \mathbf{R}_1^2) for arbitrary kites. The proof there works in this setting with only minor changes.

Here we explain a different proof. The return Lemma is obvious for points of the form (x, y) with $|x| > 20$. These points just circulate around the kite, nearly following a giant octagon before coming back to the strip. For points in the strip Σ having $|x| \leq 20$, the calculation we make in connection with the Pinwheel Lemma in §8.2 in particular establishes the Return Lemma. The calculation we make there simply involves covering $[-20, 20] \times [-2, 2]$ with 572 convex polygons such that the first return map exists and is well defined on the interior of each tile. We also prove that the remaining points, the ones in the boundaries of our tiles, do not have well defined orbits. ♠

The Return Lemma allows us to work with Ψ rather than ψ . We state the following result in terms of the action of Ψ^+ on Σ^+ . The same result holds with $(-)$ in place of $(+)$.

Lemma 4.3 *There is a canonical bijection between the unbounded ψ orbits and the unbounded Ψ^+ orbits. The bijection is such that a given Ψ orbit corresponds to the ψ orbit that contains it. Two Ψ^+ orbits are coarse equivalent if and only if the corresponding ψ orbits are coarse equivalent.*

Proof: We associate to each Ψ^+ orbit the unique ψ orbit that contains it. This injective association is also surjective, by Statement 1 of the Return Lemma. Given our description of Ψ for points far from the origin, it is clear that we can reconstruct the coarse equivalence class of a ψ -orbit from the coarse equivalence class (defined the same way) for the corresponding Ψ^+ orbit. Up to a uniformly bounded error, we obtain the graph of the ψ orbit from the graph of the Ψ^+ orbit by attaching a centrally symmetric octagon of a suitable radius to each point of Γ that is sufficiently far from the origin. ♠

4.3 An Unboundedness Criterion

Here we establish a useful criterion for the unboundedness of the Ψ orbits. Let \mathbf{Z}_+ denote the set of positive integers and let \mathbf{Z}_- denote the set of negative integers. We say that a subset $S \subset \mathbf{Z}$ is *uniformly dense* in \mathbf{Z}_+ if there is some N such that every point of \mathbf{Z}_+ is within N units of \mathbf{Z}_+ . We make the same definition relative to \mathbf{Z}_- .

Lemma 4.4 *Let O be an infinite Ψ orbit. Suppose that there is an open horizontal line segment S such that $O \cap S$ is a nonempty and nowhere dense subset of S . Then O is unbounded in both directions.*

Proof: We will assume that O is bounded in the forward direction and derive a contradiction. Let $A = \phi^{-3}$. Given the locations of the vertices of the Penrose kite, we have the following formula.

$$\Psi(p) - p = (2n_1A + 2n_2, 2n_3); \quad n_1 + n_2 + n_3 \equiv 0 \pmod{2}. \quad (36)$$

When p is far from the origin, the orbit of p stays within a uniformly thin tubular neighborhood of a centrally symmetric octagon, as we mentioned in connection with Statement 3 of the Return Lemma. The sides of this octagon, which depends on p , are always integer multiples of the vectors listed in §4.2. Moreover, opposite sides have the same length. For this reason, the vectors entering into the sum that defines $\Psi(p) - p$ nearly cancel in pairs, and we find that there is a uniform bound to $\max(|n_1|, |n_2|)$ in Equation 36.

Let (x_0, y_0) be some point of O . We can find integers (a_n, b_n) such that

$$\Psi^n(x_0, y_0) = (x_0, y_0) + 2a_nA + 2b_n. \quad (37)$$

The differences $|a_{n+1} - a_n|$ and $|b_{n+1} - b_n|$ are uniformly bounded.

The sequence $\Omega = \{a_nA + b_n\}$ is both infinite and bounded. Hence, $\Omega_1 = \{a_n\}$ has infinitely many values. Our uniform bound on $|a_{n+1} - a_n|$ now implies that Ω_1 is uniformly dense in at least one of \mathbf{Z}_+ or \mathbf{Z}_- . Combining this fact with the fact that Ω is bounded, we see that there is some N such that the union

$$\Omega^* = \bigcup_{i,j < N} \Omega + (i, j) \quad (38)$$

either contains every integer combination of the form $aA + b \in (0, 1)$ with $a > 0$, or every such integer combination with $a < 0$. In either case, Ω^* is dense in $(0, 1)$.

But Ω^* is a finite union of translates of Ω . We have shown that a finite union of translates of Ω is dense in $(0, 1)$. But the set $x_0 + 2\Omega$ is a subset of O . Hence, a finite union of translates of O is dense in some line segment. Since O is contained in the union of two lines, and the finite union of nowhere dense linear subsets is again nowhere dense, O intersects some line segment in a set that is not nowhere dense. Since the map Ψ is a piecewise translation, $O \cap S$ is not nowhere dense in S . This contradiction finishes our proof. ♠

4.4 The Arithmetic Graph

The arithmetic graph gives us a good way to visualize the orbits of the first return map $\Psi : \Sigma \rightarrow \Sigma$. Unlike our earlier papers [S1] and [S2], the arithmetic graph does not play an important role in our proofs. However, we find it very useful to illustrate some concepts with the arithmetic graph.

For $y \in (0, 2)$ we define

$$\Sigma_y = (\mathbf{R} \times \{y\}) \cup (\mathbf{R} \times \{y - 2\}) \subset \Sigma \quad (39)$$

The set Σ_y is a union of 2 horizontal lines. (The case $y = 0, 2$ leads to a trivial picture, and we ignore it.) In light of Equation 36, it makes sense to try to understand our orbits in terms of the triples of integers (n_1, n_2, n_3) rather than in terms of the values $(2n_1A + 2n_2, n_3)$. The parity of $n_1 + n_2$ determines n_3 . Thus, the pair of integers (n_1, n_2) really determines the behavior of Ψ on p .

Fixing some $y \in (0, 2)$, we define a map $M_y : \mathbf{Z}^2 \rightarrow \Sigma_y$ as follows. We first choose some *offset value* $\xi \in \mathbf{R}$. This offset value selects which orbits we focus on. Given a point $(m, n) \in \mathbf{Z}^2$ we define

$$M_{y,\xi}(m, n) = (2mA + 2n + \xi, y + \tau(m, n)). \quad (40)$$

Here $\tau(m, n) = 0$ if $m + n$ is odd and $\tau(m, n) = -1$ if $m + n$ is even.

It is a consequence of Equation 36 that Ψ preserves the image $M_{y,\xi}(\mathbf{Z}^2)$. That is, given $(m_0, n_0) \in \mathbf{Z}^2$, there is another $(m_1, n_1) \in \mathbf{Z}^2$ such that $\Psi \circ M(m_0, n_0) = M(m_1, n_1)$. We write $(m_0, n_0) \rightarrow (m_1, n_1)$ in this case. Here we have set $M = M_{y,\xi}$.

Given (y, ξ) , we form the arithmetic graph $\Gamma(y, \xi)$ as follows: We include (m, n) as a vertex of Γ if and only if $M(m, n)$ has a well defined orbit. We connect the vertex (m_0, n_0) to the vertex (m_1, n_1) if and only if $(m_0, n_0) \rightarrow$

(m_1, n_1) . In this way, we produce a 2-valent directed graph in the plane, whose vertices lie in \mathbf{Z}^2 . For kites, the graph Γ is always embedded. We proved this for the special orbits (i.e. those on Σ_1) relative to any kite parameter in [S2]. The proof for the general orbit is similar, though we have not written down the details.

Figure 4.3 shows a portion of the unique unbounded γ component of the graph $\Gamma(1, \phi^{-2})$. (The straight line segment at the bottom is just for reference.) The straight line is $M^{-1}(0)$. In [S1] we studied this component and showed that indeed γ is unbounded.

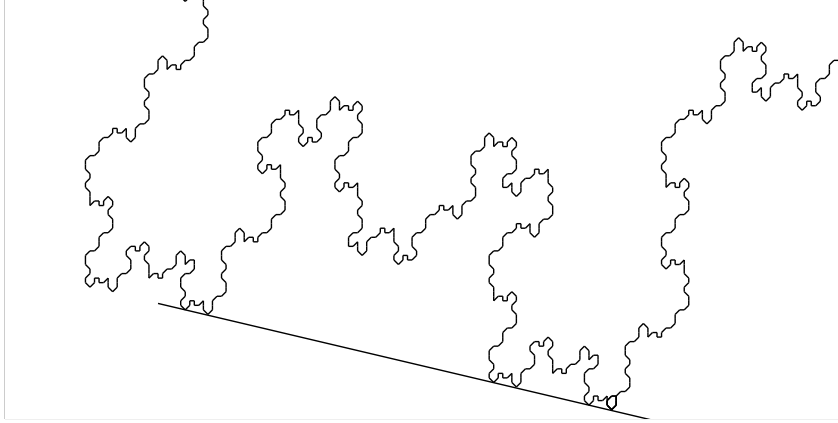


Figure 4.3: The unbounded component of $\Gamma(1, \phi^{-2})$.

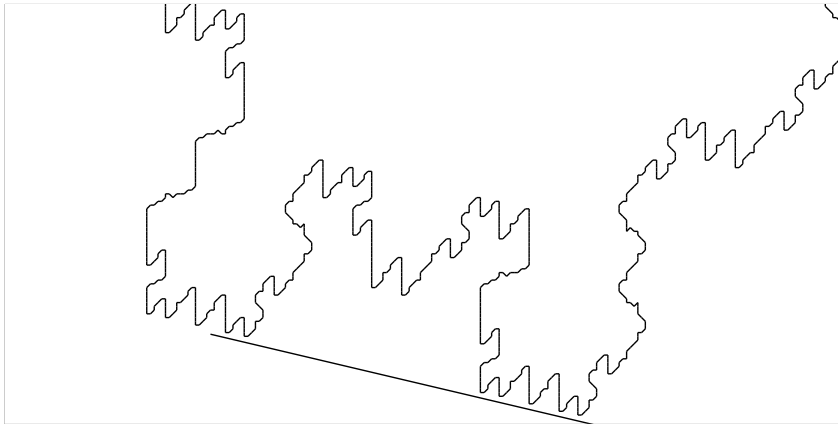


Figure 4.4: The unbounded component of $\Gamma(\phi^{-3}, \phi^{-5})$.

Figure 4.4 shows the unbounded component of $\Gamma(\phi^{-3}, \phi^{-5})$. This component is quasi-isometric to the one in the previous picture but looks different locally. This is an illustration of Theorem 1.5 in action.

4.5 The Freezing Phenomenon

The arithmetic graph illustrates an interesting phenomenon. Figure 4.5 shows the portion of a graph that corresponds to the parameter $y = \phi^{-5}$. Notice the long range linear order. This phenomenon becomes more and more extreme as $y \rightarrow 0$: Longer and longer portions of the arithmetic graph contain these nearly linear portions. The components of the arithmetic graph follow along the lines, switching onto a new line at every intersection so as to avoid collisions. We call this the *freezing phenomenon* because the orbits seem to freeze into a characteristic shape.

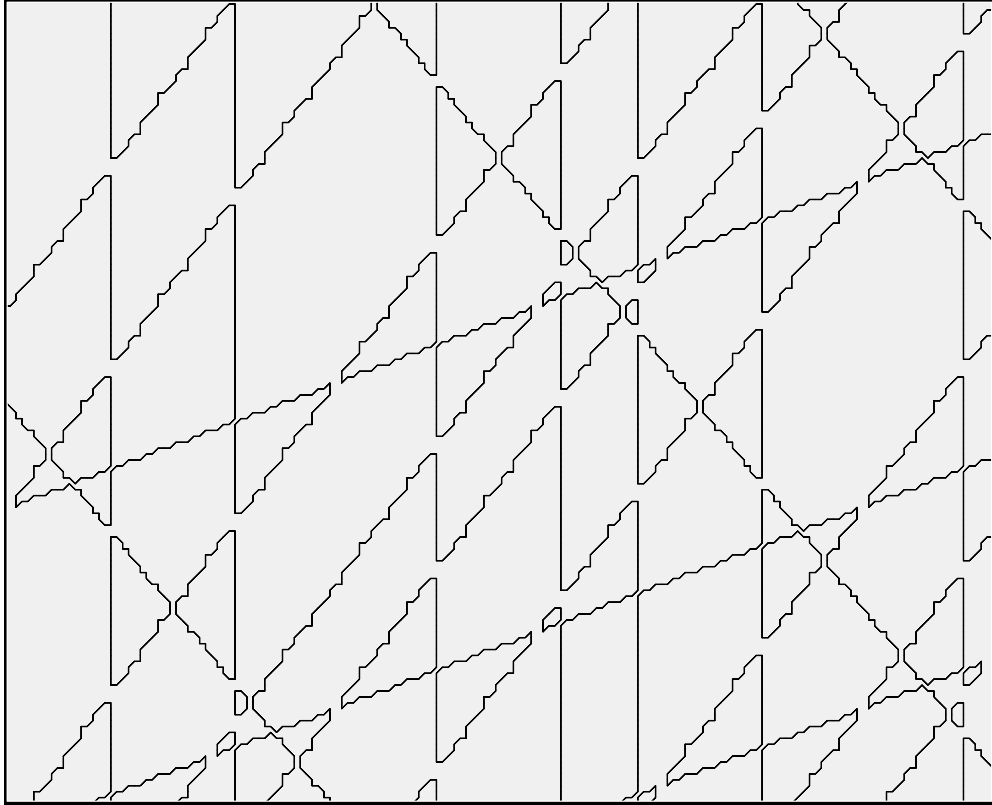


Figure 4.5: Part of $\Gamma(\phi^{-5}, -2\phi^{-4})$.

The freezing phenomenon seems almost to be in contradiction with Theorem 1.5, as we now explain. Let R be the renormalization map. The map R^2 is expanding, so the full preimage of any point $y \in \mathbf{R}/2\mathbf{Z}$ is dense. Hence, Theorem 1.5 has the following corollary.

Corollary 4.5 *Let O be any generic unbounded orbit. Then there is a dense set $y \in (0, 2)$ such that U_y contains an orbit that is both coarsely and locally similar to O . In particular, this is true for a sequence $\{y_n\}$ converging to 0.*

What makes Corollary 4.5 seem at odds with the freezing phenomenon is that the arithmetic graph is definitely changing shape as $y \rightarrow 0$, but somehow the global shape is always coarsely equivalent to some fixed arithmetic graph. The escape from the contradiction is that the bi-lipschitz constant implicit in the definition of coarse equivalence tends to ∞ as y tends to 0.

Given that both the freezing phenomenon and Corollary 4.5 are true, the renormalization discussed in our theorems must be expressible directly in terms of the multigrid system of lines. This is indeed the case. On the level of the multigrid, the renormalization is reminiscent both of the renormalization one sees for Sturmian sequences and the renormalization one sees for the Penrose tilings, especially when it is expressed in terms of De Bruijn's pentagrids. See [DeB]. So, there really is an underlying connection between outer billiards on the Penrose kite and the Penrose tiling.

The freezing phenomenon works for all kites, and it is part of a larger phenomenon, though I don't see the renormalization scheme for a general kite. In [S3] I wrote some informal notes describing the connection between the arithmetic graph and the multigrids of lines, but I did not include any proofs. I don't currently know any.

5 Structural Results

5.1 Compactification

The map $\Psi : \Sigma \rightarrow \Sigma$ turns out to be an infinite golden polygon exchange map. This quasi-periodicity is the driving idea behind our next result. Define the flat 3-torus

$$\widehat{\Sigma} = \mathbf{T}^3; \quad \mathbf{T} = \mathbf{R}/2\mathbf{Z}. \quad (41)$$

Theorem 5.1 (Compactification) *There is a fibered golden polyhedron exchange map $\widehat{\Psi} : \widehat{\Sigma} \rightarrow \widehat{\Sigma}$, and an injective embedding $\Theta : \Sigma \rightarrow \widehat{\Sigma}$, given by the equation*

$$\Theta(x, y) = \left(1, \frac{1}{2}, 0\right) + \left(\frac{x}{\phi}, \frac{x-y}{2}, y\right) \quad (42)$$

which is a semi-conjugacy between Ψ and $\widehat{\Psi}$.

Some terms require explanation. First, the coordinates of Θ are interpreted as living in \mathbf{T} . The translational part of Θ is somewhat arbitrary. We found this choice convenient for the purpose of drawing pictures. To say that Θ is a *semi-conjugacy* is to say that $\widehat{\Psi} \circ \Theta = \Theta \circ \Psi$ wherever all maps are defined. Since Θ is injective, we will often identify Σ with the subset $\Theta(\Sigma) \subset \widehat{\Sigma}$. In this way, we think of $\widehat{\Sigma}$ as a compactification of Σ . With this interpretation, the semi-conjugacy condition just says that $\widehat{\Psi}$ extends Ψ .

The set $\Theta(\Sigma)$ is contained in a countable dense union of parallel planes that are transverse to the horizontal planes. We define

$$\Sigma_y = \left(\mathbf{R} \times \{y\}\right) \cup \left(\mathbf{R} \times \{y-2\}\right). \quad (43)$$

for $y \in (0, 2)$. Given that ψ preserves the set $\mathbf{R} \times (y + 2\mathbf{Z})$, the map Ψ preserves Σ_y for each y . The image $\Theta(\Sigma_y)$ is densely contained in a single horizontal plane, and Θ is a semi-conjugacy between the map $\Theta : \Sigma_y \rightarrow \Sigma_y$ and a exchange map on the corresponding horizontal plane.

The semi-conjugacy between $\Psi : \Sigma_1 \rightarrow \Sigma_1$ and $\widehat{\Psi} : \widehat{\Sigma}_1 \rightarrow \widehat{\Sigma}_1$ is equivalent to the Arithmetic Graph Lemma of [S1]. Here $\widehat{\Sigma}_1$ is the horizontal plane of height 1. Similarly, Theorem 5.1 is closely related to the Master Picture Theorem in [S2]. See §7 for a discussion.

5.2 Structure of the Compactification

In terms of raw data, we describe the compactification precisely in §14. Here we highlight the important features. We begin by showing pictures. Each picture is a different horizontal slice of $\widehat{\Sigma}$. Let R be the parallelogram with vertices

$$(-A, 0); \quad (A, 2); \quad (-A+2, 0); \quad (A+2, 0); \quad A = \phi^{-3} = \sqrt{5}-2. \quad (44)$$

The solid $X = R \times [0, 2]$ is a fundamental domain for the action of $2\mathbf{Z}^3$ on \mathbf{R}^3 and conveniently we can choose lifts of our polyhedra so that they give a partition of X . Figure 5.1 shows the intersection of X with the plane $z = 0$. The grey polygons represent the intersections of the polyhedra in our partition with this plane.

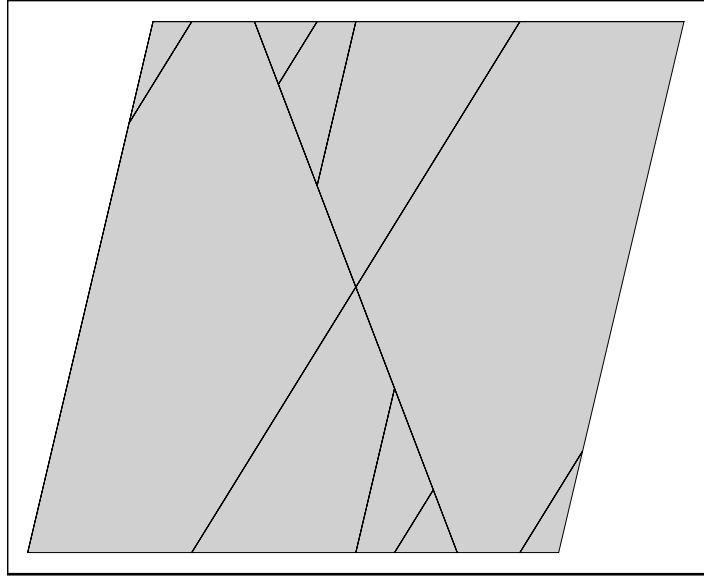


Figure 5.1: The slice at 0.

Below we show the slices at the heights

$$0; \quad \phi^{-3}; \quad \phi^{-2} \quad \phi^{-1} \quad 2\phi^{-2}; \quad 1. \quad (45)$$

The slices at heights t and $2 - t$ are isometric to each other, via the isometry which rotates 180 degrees about the midpoint of the parallelogram. Indeed, the map $(x, y, z) \rightarrow (2 - x, 2 - y, 2 - z)$ is an automorphism of the partition. The slices values we choose are precisely the values in $[0, 1]$ which contain some polyhedron vertices. Here are the remaining slices

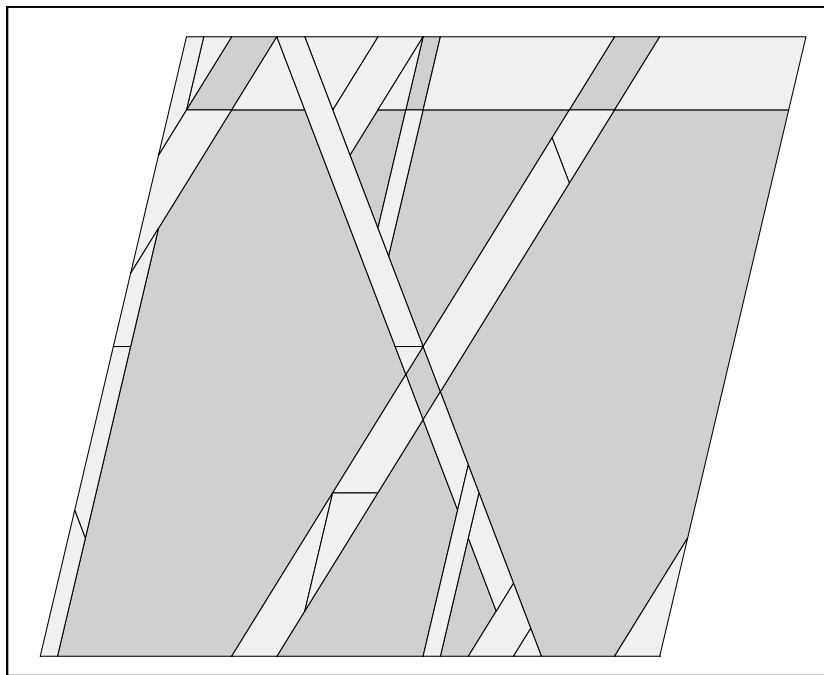


Figure 5.2: The slice at ϕ^{-3} .

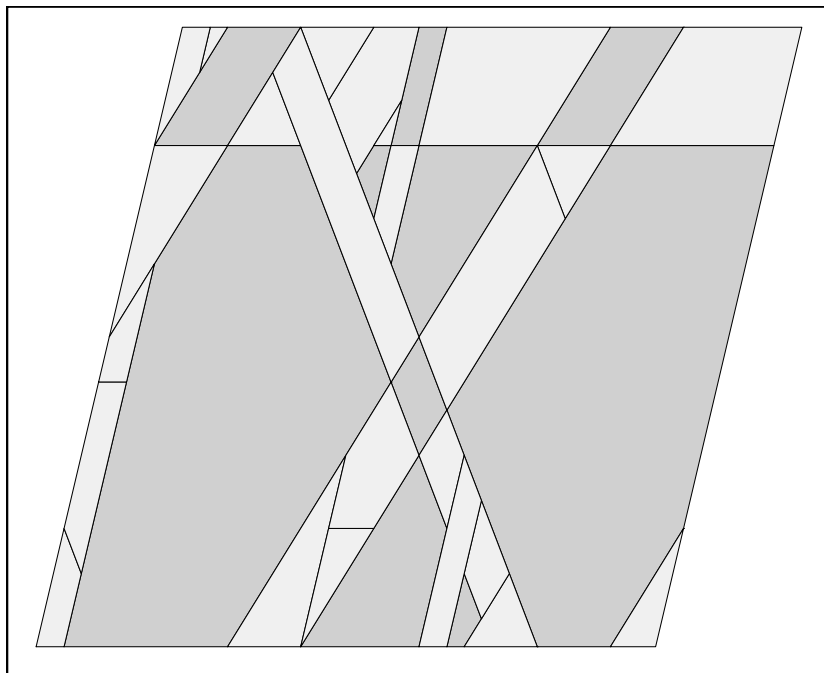


Figure 5.3: The slice at ϕ^{-2} .

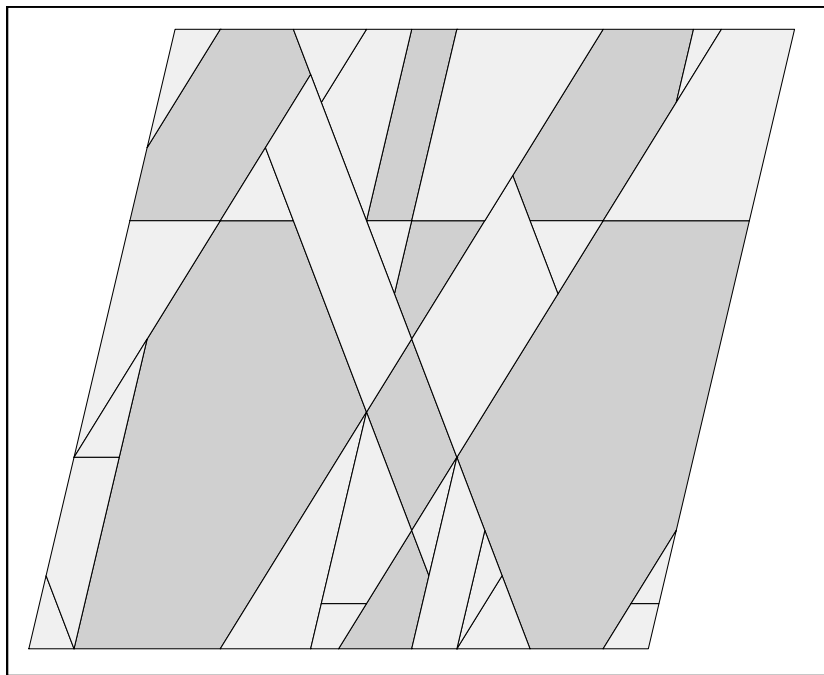


Figure 5.4: The slice at ϕ^{-1} .

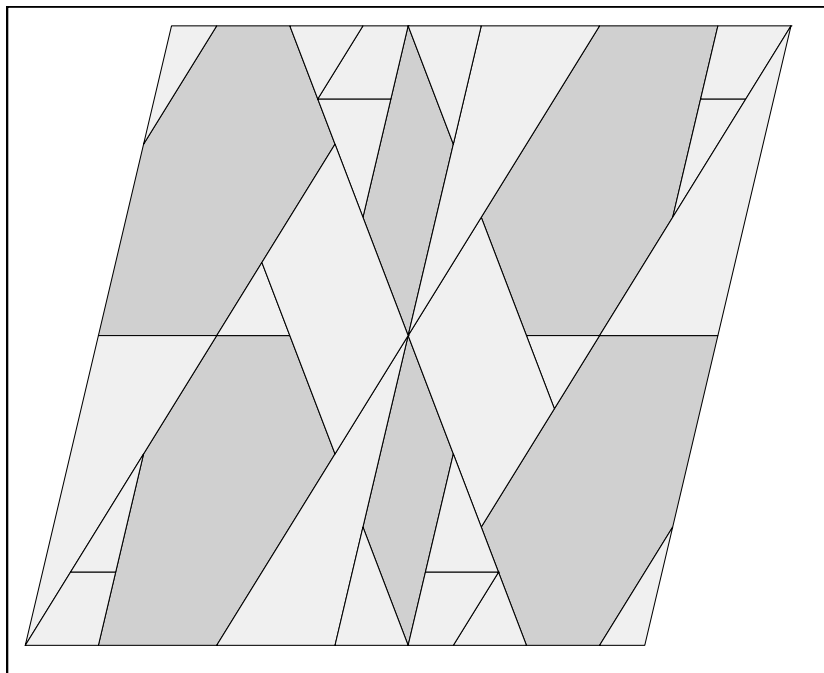


Figure 5.5: The slice at $2\phi^{-2}$.

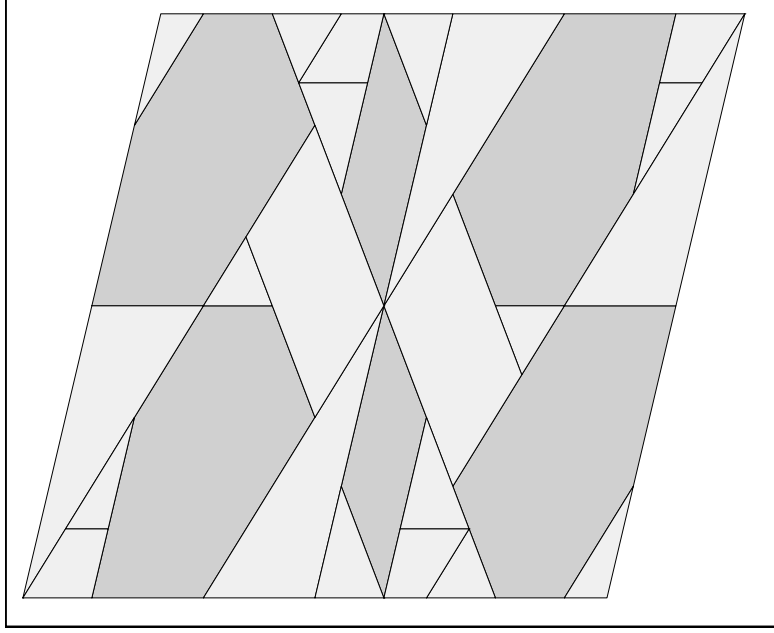


Figure 5.6: The slice at 1.

The shading in our pictures has the following explanation. The map $\widehat{\Psi}$ is the identity on a polygon iff the polygon is colored grey. In particular, $\widehat{\Psi}$ is the identity on the slice $z = [0]$.

Now we discuss the action of $\widehat{\Psi}$ on the partition. Say that a *special affine involution* of $\widetilde{\Sigma}$ is an order 2 affine map that preserves each horizontal slice and is an isometric rotation in each horizontal slice. The map

$$\tau(x, y, z) \rightarrow (-x, -3z - y, z) \quad (46)$$

is one such map. For each polyhedron P in the partition, there is a special affine involution I_P such that $I_P(P)$ is also in the partition. We have

$$\widehat{\Psi}|_P = \tau \circ \iota_P. \quad (47)$$

Lemma 5.2 $\widehat{\Psi}$ is a fibered golden polyhedron exchange map, as in §4.1.

Proof: A direct calculation shows that every edge of every polyhedron in the partition satisfies Equation 33. The equations for the vertices of these polyhedra are listed in the appendix. ♠

There is one other feature of the pictures we would like to mention. The lines in Figure 5.1 are present in all the pictures. These “persistent lines” divide $\widehat{\Sigma}$ into 10 convex prisms – i.e. polyhedra of the form $P \times [0, 2]$, where P is a convex polygon in the plane. Hence, the 64 polyhedra in the partition can be divided into 10 groups, each of which partitions one of the prisms. Figure 5.7 illustrates this for the slice at height 1. Our java program allows the user to highlight the clusters one at a time.

In particular, the 64 polyhedra together partition the fundamental domain \widehat{F} , and not just the torus. In other words, the outer edges in all the figures are actually part of the polyhedra, and not just artifacts of the way we have drawn the picture. A moment’s reflection reveals that this issue arises whenever one wants to depict a partition of a torus by polygons.

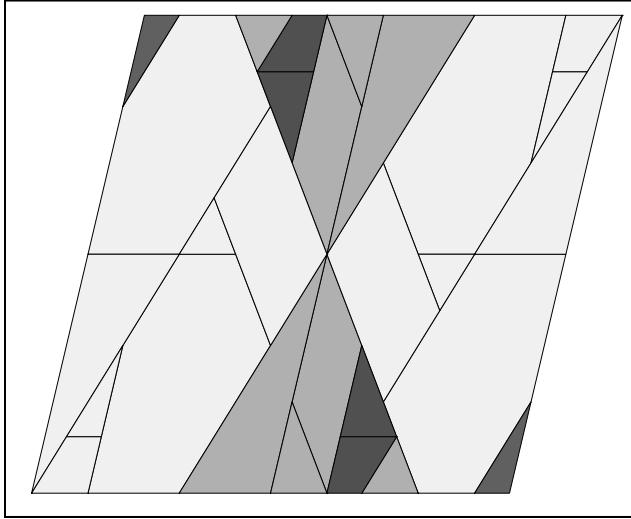


Figure 5.7: The slice at 1, divided into prisms.

There is a second way to group the 64 polyhedra into 10 (non-vertical) prisms. If P is one of the original 10 prisms, then $\tau(P)$ is another prism which is a finite union of some of the pieces. Here τ is the affine involution defined above. Thus, the partition of $\widehat{\Sigma}$ into 64 convex polyhedra is compatible with two interlocking partitions of $\widehat{\Sigma}$ into 10 prisms each.

The existence of the two families of prisms suggests that our polyhedron exchange map is actually the square of a piecewise affine map that is defined in terms of the two prism partitions. This is indeed the case, but we did not find this characterization useful.

5.3 The Renormalization Theorem

We call an open set $\hat{A} \subset \hat{\Sigma}$ *atomic* if \hat{A} has a finite partition into golden polyhedra $\hat{P} \cup \dots \cup \hat{P}_n$ such that the first return map $\Psi : \hat{A} \rightarrow \hat{A}$ is entirely defined, and a translation, when restricted to the interior of each \hat{P}_k . We call the polyhedra \hat{P}_k the *atoms* of \hat{A} . Our definition does not uniquely define the atoms, but in practice our atoms will be the maximal ones.

Suppose that \hat{A} and \hat{B} are two atomic sets. We say that a map $f : \hat{A} \rightarrow \hat{B}$ is an *atomic bijection* if f bijectively maps the atoms of \hat{A} to the atoms of \hat{B} , and the restriction of f to each atom is a homothety.

We say that a *layer* of \hat{A} is the set of all points in \hat{A} whose third coordinates belong to some interval. We call $f : \hat{A} \rightarrow \hat{B}$ an *atomic cover* if \hat{A} has a partition into layers $\hat{A}_1, \dots, \hat{A}_{dk}$ and \hat{B} has a partition into slabs $\hat{B}_1, \dots, \hat{B}_k$ such that f is surjective and d -to-1 and $f : \hat{A}_i \rightarrow \hat{B}_j$ is an atomic bijection for each i . Here the index j depends on i . Each index j corresponds to d indices i . Assuming that the map f is given, we call the abovementioned layers *the layers* of \hat{A} and \hat{B} .

An orbit (or orbit portion) is *generic* if it does not intersect any non-horizontal plane that is defined over $\mathbf{Z}[\phi]$.

Theorem 5.3 (Renormalization) *There is a pair of atomic sets $\hat{A}, \hat{B} \subset \Sigma$ and a 3-to-1 atomic covering map $\hat{R} : \hat{A} \rightarrow \hat{B}$ with the following properties.*

1. \hat{R} conjugates $\hat{\Psi}|_{\hat{A}}$ to $\hat{\Psi}|_{\hat{B}}$ or to $\hat{\Psi}^{-1}|_{\hat{B}}$, according as \hat{R} acts as a translation or a dilation.
2. \hat{R} acts on the horizontal planes exactly as the map R acts on $\mathbf{R}/2\mathbf{Z}$.
3. \hat{R} maps $\Theta(\Sigma) \cap \hat{A}$ to $\Theta(\Sigma) \cap \hat{B}$.
4. Any generic orbit portion of length 812 intersects \hat{A} and any generic orbit portion of length 109 intersects \hat{B} .
5. For any generic $p \in \hat{A}$, the orbit of $\hat{R}(p) \in \hat{B}$ returns to \hat{B} in fewer steps than the orbit of p returns to \hat{A} .
6. Any generic orbit which intersects \hat{A} also intersects \hat{B} .

Remark: The constant 109 is optimal. The constant $812 = 703 + 109$ is an artifact of our proof. The smaller constant 703 would be optimal. See §10.5 for a discussion of these matters.

5.4 Structure of the Renormalization

In terms of raw data, we describe \hat{A} , \hat{B} , and \hat{R} precisely in §14.3 and §14.4. In this section we give the reader a feel for these objects.

5.4.1 The Set \hat{B}

We describe \hat{B} precisely in §14.3.

The partition

$$0 < \phi^{-2} < 2\phi^{-2} < 1 < 2 - \phi^{-2} < 2 - 2\phi^{-2} < 2. \quad (48)$$

divides $\mathbf{R}/2\mathbf{Z}$ into 6 intervals, $J_1, J_2, J_{31}, J_{32}, J_4, J_5$. This is the same partition that defines the intervals I_1, \dots, I_5 , except that the value 1 has been inserted, so as to split I_3 into two symmetric halves. The layers of \hat{B} (relative to \hat{R}) are the ones corresponding to these intervals.

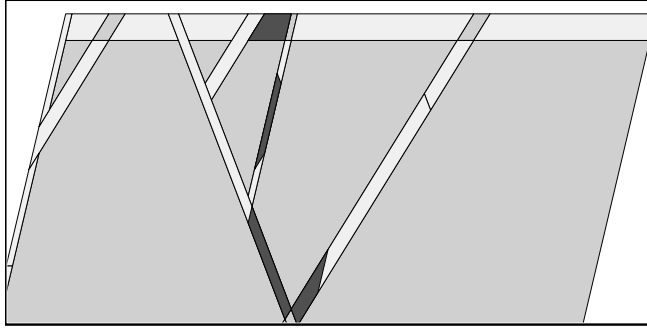


Figure 5.8: The slice of \hat{B} at ϕ^{-6} .

Figure 5.8 shows a slice of \hat{B} at the parameter $z = \phi^{-6}$. The 4 dark polygons compose the slice. We are showing the top half of $\hat{\Sigma}_z$. This slice is part of the first layer \hat{B}_0 . Each layer \hat{B}_i decomposes into 4 convex polyhedra, which we call *branches*. Each branch further decomposes into between 9 and 48 atoms on which the first return map is well defined.

The fact that each layer of \hat{B} decomposes into 4 branches is practically forced by the structure $\hat{\Psi}$, as we now explain. For z near 0, the restriction of $\hat{\Psi}$ to the slice $\hat{\Sigma}_z$ is very nearly the identity. It fails to be the identity only in a thin neighborhood of 10 line segments. There is a set of 4 slopes such that each line segment has one of these 4 slopes. In this way, the 10 “active strips” of $\hat{\Sigma}_z$ are nearly partitioned into 4 groups. (We say *nearly* because these strips intersect.)

As $z \rightarrow 0$ there exist arbitrarily long orbits which remain within a single group. If we want to choose \hat{B} so that every sufficiently long orbit intersects \hat{B} , we must include a polyhedron that nontrivially intersects each of the 4 groups. The choice of these polyhedra is not uniquely determined by this requirement, but we have made a reasonable and efficient choice. Once we make our choices for z near zero, the picture is determined for the entire interval $[0, \phi^{-2}]$ by a kind of continuation principle: The polyhedra naturally open up and follow along as the active strips get wider.

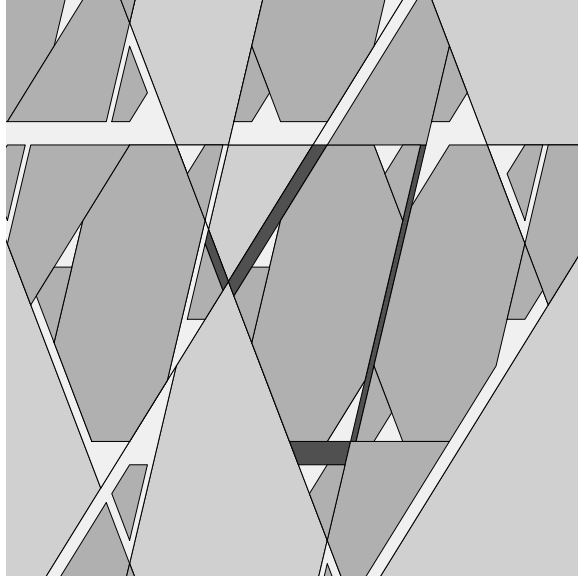


Figure 5.9: The slice of \hat{B} (dark grey) at $-17 + 11\phi$.

We say that the *freezing phenomenon* is the tendency of the long orbits to align along thin strips. This phenomenon is another manifestation of the one discussed in §4.5. What happens in the first interval happens in each of the 6 intervals. The parameters $2\phi^{-2}$ and $2 - 2\phi^{-2}$ also exhibit the freezing phenomenon. Figure 5.9 shows the slice of \hat{B} at the parameter $-17 + 11\phi$. This parameter is fairly near $2\phi^{-2}$. The light grey polygons are some periodic tiles. These periodic tiles nearly fill up the slice, but they leave some thin cracks. These thin cracks line up along the 4 basic directions, and the long orbits accumulate in the cracks. The set \hat{B} fits inside the cracks, with one polyhedron per direction. We will list the vertices of the polyhedra comprising \hat{B} in the appendix. The reader can see much better pictures of \hat{B} (and \hat{A}) using our applet.

5.4.2 The set \hat{A}

The partition associated to \hat{A} consists of the 18 intervals we get by pulling back the J partition under the action of R . Precisely:

- Within I_1 , the partition is $R^{-1}(J_1), \dots, R^{-1}(J_4)$.
- Within I_2 , the partition is $R^{-1}(J_5)$.
- Within I_3 , the partition is $R^{-1}(J_1), \dots, R^{-1}(J_5)$.
- Within I_4 , the partition is $R^{-1}(J_1)$.
- Within I_5 , the partition is $R^{-1}(J_2), \dots, R^{-1}(J_5)$.

The 18 layers of \hat{A} correspond to these 18 intervals. Again, each natural piece is decomposed into 4. The map \hat{R} carries each branch of each layer of \hat{A} homothetically to the corresponding branch of the corresponding layer of \hat{B} . As with the \hat{B} -branches, the \hat{A} -branches are further decomposed into atoms.

5.4.3 The Map \hat{R}

Here we highlight the general features. Let \hat{R}_{ijk} denote the homothety that expands distances by ϕ^3 and fixes the point (i, j, k) .

1. When restricted to $\hat{A}(I_1)$, the map is \hat{R}_{110} or \hat{R}_{120} .
2. When restricted to $\hat{A}(I_2)$, the map is translation by $(0, 0, 2\phi^{-1})$.
3. When restricted to $\hat{A}(I_3)$, the map is one of R_{1k1} for $k \in \{0, 1, 2\}$.
4. When restricted to $\hat{A}(I_4)$, the map is translation by $(0, 0, -2\phi^{-1})$.
5. When restricted to $\hat{A}(I_5)$, the map is one of \hat{R}_{102} or \hat{R}_{112} .

Now we will show the Renormalization Theorem in action. Figure 5.10 shows a closeup of the set \hat{A} at the slice $5 - 3\phi$ together with part of an orbit.

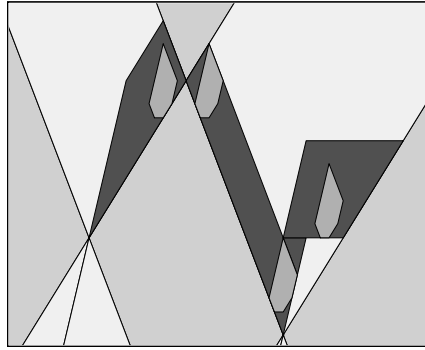


Figure 5.10: Closeup of an orbit.

Figure 5.11 zooms out, to reveal much more of the orbit in the same slice. The set \hat{A} is at the very bottom.

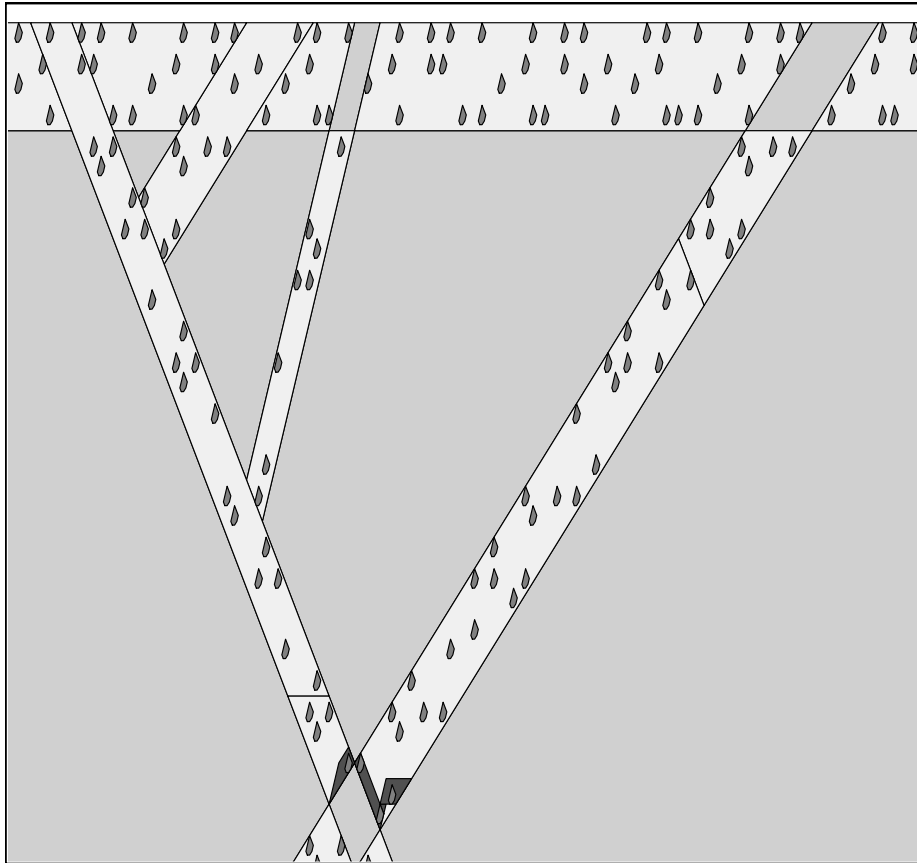


Figure 5.11: The slice of \hat{A} at $5 - 3\phi$, together with an orbit.

Now for the magic trick. Figure 5.12 shows the slice of \widehat{B} at the parameter $-1 + \phi = R(5 - 3\phi)$, together with an orbit. The orbit in Figure 5.12 looks different overall from the orbit in Figure 5.11, but the new orbit intersects \widehat{B} in the same way that the old orbit intersects \widehat{A} . In other words, were we to take a closeup of just the renormalization set, we couldn't tell whether it was the \widehat{A} -set inside the first slice or the \widehat{B} -set in the second slice. The reader can see many more pictures like this using our applet.

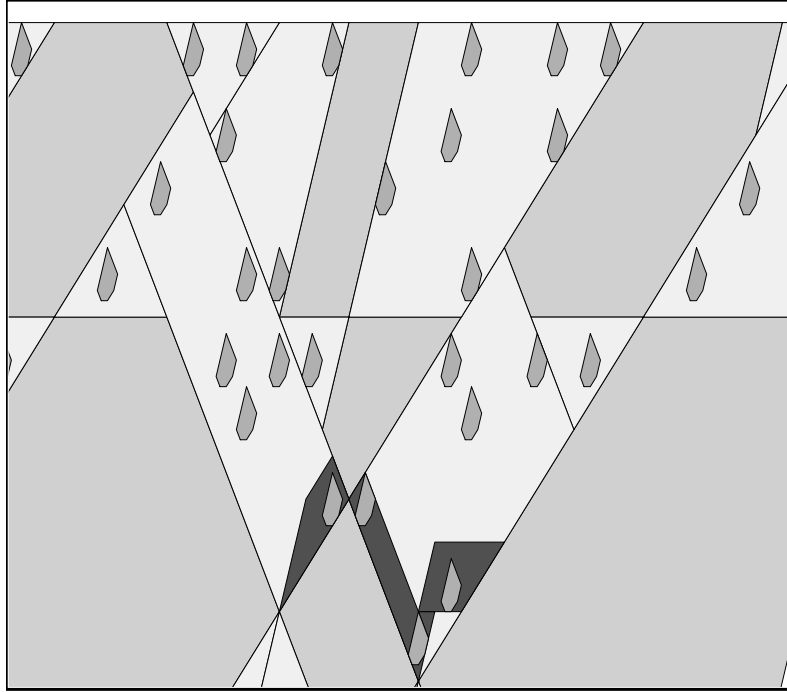


Figure 5.12: The slice of \widehat{B} at $-1 + \phi$, together with an orbit.

The Renormalization Theorem and the freezing phenomenon together combine to give a nice qualitative description of the dynamical tiling associated to the exchange map. Consider a parameter z_0 such that all orbits are periodic. (Any parameter z_0 such that $R^n(z_0) = 0$ will have this property.) At z_0 , the slice is tiled by periodic tiles. As the parameter moves away from z_0 , some cracks open up, and these cracks are filled with smaller periodic tiles associated to longer periodic orbits. If the new parameter is chosen just right, the slice is again tiled by periodic tiles: the original large ones and the much smaller ones that live in the cracks. Moving the parameter a bit, some new cracks open up and even smaller tiles fill in these cracks. And so on.

5.5 Renormalization of Orbits

Genericity: We call a point of $\widehat{\Sigma}$ *generic* if it does not lie in any non-horizontal plane that is defined over $\mathbf{Z}[\phi]$. The map Θ carries the set of generic points in Σ into the set of generic points of $\widehat{\Sigma}$. Moreover, the translation vectors defining $\widehat{\Psi}$ all lie in $\mathbf{Z}[\phi]^3$. Hence $\widehat{\Psi}$ preserves the set of generic orbits. All the atoms of \widehat{A} and \widehat{B} have their faces in the planes we have excluded. Hence, any generic point in \widehat{A} or \widehat{B} lies in the interior of an atom. See §5.8 below for a further discussion.

Basic Definition: The Renormalization Theorem allows us to define a *renormalization operation* on generic orbits. Let O_1 be a generic orbit that intersects \widehat{A} . We know that O_1 does not intersect the boundaries of any of the atoms. We start with $p_1 \in O_1 \cap \widehat{A}$. We let $p_2 = \widehat{R}(p_1)$ and we let O_2 be the orbit of p_2 . This definition is independent of choice of p_1 thanks to Item 1 of the Renormalization Theorem. For instance, if we choose the first return point $p'_1 = \Psi^k(p_1)$ for a suitable power of k , then $\widehat{R}(p'_1) = p'_2 = \widehat{\Psi}^{\pm k}(p_2)$. We write

$$O_1 \rightsquigarrow O_2 \tag{49}$$

when O_1 and O_2 are related as above. We call O_2 the *renormalization* of O_1 . Given the nature of the map \widehat{R} , the orbit O_2 is also generic.

Distinguished Orbits: There is one situation where our construction above is ambiguous. We call an orbit \widehat{A} -distinguished if it lies in a horizontal plane that contains the top or bottom of one of the \widehat{A} layers. These orbits are somewhat of a nuisance. When O_1 is a distinguished orbit, we define the operation $O_1 \rightsquigarrow O_2$ by including O_1 in one layer of \widehat{A} or the other. We will not take the trouble to prove the O_2 does not depend on the choice of layer because in the cases of interest to us, the uniqueness either doesn't matter or comes as a byproduct of our proof.

Renormalization of Cores: We now mention a more precise kind of renormalization. Say that a \widehat{A} -core is an orbit portion p_1, \dots, p_n such that p_1 and p_{n+1} lie in \widehat{A} but p_2, \dots, p_n do not. We define \widehat{B} -cores in the same way. If α is an \widehat{A} -core and β is a \widehat{B} -core, we write $\alpha \rightsquigarrow \beta$ if \widehat{R} maps the endpoints of α to the endpoints of β . In this case, we have $O_1 \rightsquigarrow O_2$, where O_1 is the orbit containing the points of α and O_2 is the orbit containing the points of β .

Restriction to the Strip: So far we have been talking about the picture in the compactification $\widehat{\Sigma}$, but we can transfer everything over to Σ . By the Compactification Theorem, we can consider Σ as a subset of $\widehat{\Sigma}$. With this interpretation, the action of Ψ is just the restriction of $\widehat{\Psi}$. By Item 4 of the Renormalization Theorem, the operation $O_1 \rightsquigarrow O_2$ preserves the set of generic infinite Ψ -orbits.

Coarse Equivalence: The arithmetic graph illustrates the nature of the renormalization map on cores. We say that a *strand* is the arithmetic graph of a core, translated so that one endpoint is the origin. The renormalization operation on cores gives a map from the set of all A -strands to the set of all B -strands. Figure 5.13 plots an A -strand in black and the corresponding B -strand in grey. The two strands start at the origin, which is the endpoint at right, and they have the other endpoint in common as well. This example is what we will call type-1. It corresponds to a layer of \widehat{A} on which \widehat{R} is an isometry.

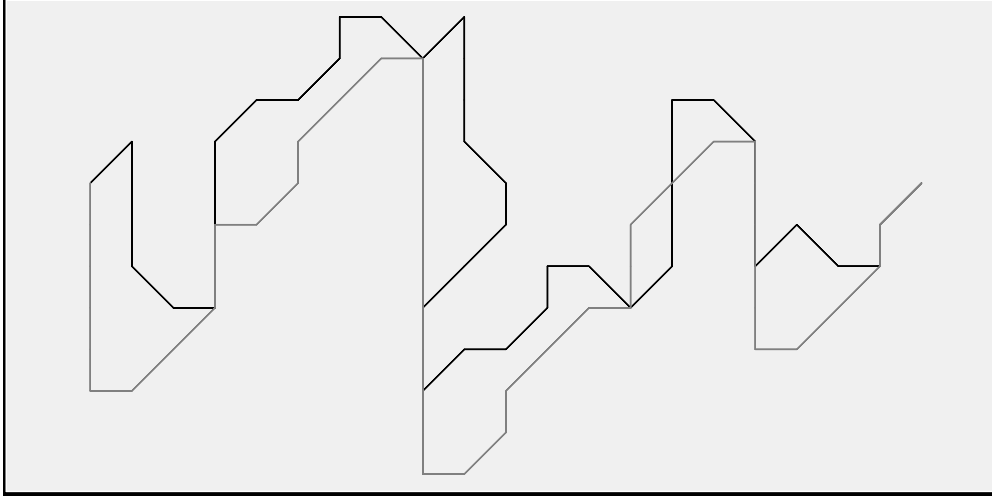


Figure 5.13: Corresponding A and B strands.

Figure 5.14 shows another example. This example corresponds to a layer of \widehat{A} where \widehat{R} is a dilation. In this example the origin is at bottom right, and we have shrunk the A -strand by a factor of ϕ^{-3} . Notice the remarkable agreement. This remarkable agreement is the subject of the Far Reduction Theorem below, and also the basis for Theorem 1.5.

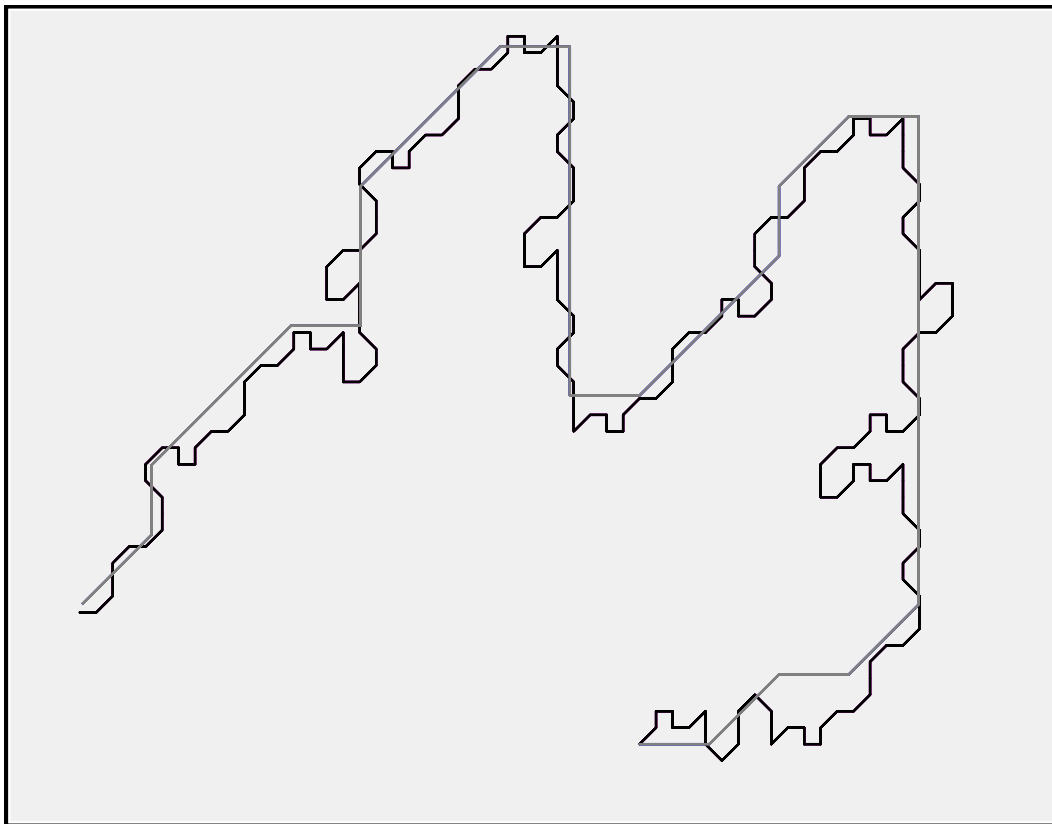


Figure 5.14: Corresponding A and B strands.

There is a total of $2034 = 3 \times 678$ A -strands and a total of 678 B -strands. Using our program, the tireless reader can see pictures like the ones above for every pair of corresponding strands.

5.6 The Fundamental Orbit Theorem

To state the Fundamental Orbit Theorem, we use the notation established in §3.4. Let T be the fundamental triangle, the subject of Theorem 1.4. Recall that \mathcal{T} is the tiling of T discussed in §3. Recall that J_0 and K_0 are the largest octagon and kite in \mathcal{T} . Say that these tiles have *depth zero*. Say that a tile of \mathcal{T} has *depth n* if it is similar to J_0 or K_0 by a factor of ϕ^{-3n} . Let $\mathcal{T}(n)$ denote the finite union of tiles having depth n or less.

Recall that ψ is the outer billiards map. It is easy to check that ψ^{-1} is entirely defined on the interior of $T^+ \cup T^-$.

Theorem 5.4 (Fundamental Orbit) *Let $p_1 \in T_{ij}^\pm - \mathcal{T}(2)$ be a generic point and let*

$$p_2 = \psi^{-1}(R_{ij}^\pm(p_1))$$

Let α_1 be an A -core that contains p_1 . Then $\alpha_1 \rightsquigarrow \alpha_2$, where α_2 is the B -core containing p_2 . In particular $O_1 \rightsquigarrow O_2$, where O_k is the orbit containing p_k .

The exponents (\pm) are written loosely in the Fundamental Orbit Theorem. The choice of $(+)$ or $(-)$ in every cases is taken so that the relevant maps have the proper domain and range.

The fundamental orbit says that, on the level of orbits, the maps R_{ij}^\pm and the renormalization map have the same action. The Fundamental Orbit Theorem is the result behind Theorem 1.4.

5.7 The Reduction Theorems

Here we present the two results which explain the qualitative action of the action of \widehat{R} on orbits.

We call two orbits O_1 and O_2 *associates* if one of 4 things holds.

- $O_1 = O_2$.
- $O_1 = \overline{O_2}$, the complex conjugate orbit.
- $O_1 = \psi'(O_2)$. Here ψ' is the outer billiards map.
- $O_1 = \psi'(\overline{O_2})$.

Associate orbits are clearly locally and coarsely similar. equivalent.

We write $O_1 \rightarrow O_2$ if $O'_1 \rightsquigarrow O'_2$, where O'_k is an associate of O_k . The two relations (\rightsquigarrow) and (\rightarrow) are practically the same. We use the latter because it allows us to prove our results with significantly less computation.

Define

$$\Sigma_{24} = [-24, 24] \times [-2, 2] \tag{50}$$

Theorem 5.5 (Near Reduction) *Let O_1 be any generic infinite orbit that intersects Σ_{24} . Then there is some m such that $O_1 \rightarrow \dots \rightarrow O_m$, and O_m intersects T , the fundamental triangle. More precisely, let $y \in (0, 2)$ be any point such that $\{R^n(y)\}$ does not contain 0. Then there is some N with the following property. If O_1 contains a point within $1/N$ of the segment $[-24, 24] \times \{y\}$, then $O_1 \rightarrow \dots \rightarrow O_m$ and O_m intersects T^+ and $m < N$.*

Given an A -core α , we define

$$|\alpha|_x = \min_{(x,y) \in \alpha} |x|. \quad (51)$$

That is, $|\alpha_x|$ measures how close the x -coordinates of α come to 0. We make the same definition for β .

We say that α has type-1 if the renormalization map \widehat{R} is a piecewise translation on the layer containing α . We say that α has type-2 if \widehat{R} is a piecewise dilation by ϕ^3 on the layer containing α .

Theorem 5.6 (Far Reduction) *Let α and β be \widehat{A} and \widehat{B} strands respectively. Suppose that $\alpha \rightsquigarrow \beta$. Then the following is true.*

1. *If α has type 1 then $|\alpha_x| - C < |\beta|_x < |\alpha|_x + 12$.*
2. *If α has type 2 then $\phi^{-3}|\alpha_x| - C < |\beta|_x < \phi^{-3}|\alpha|_x + 15$.*

Here C is a universal constant that we don't care about.

We think of the Near Reduction Theorem and the Far Reduction Theorem as geometric versions of our Descent Lemma II. The Far Reduction Theorem is similar to the theoretical argument we gave in order to reduce the Descent Lemma II to a computer calculation, and the Near Reduction Theorem is similar to this computer calculation. Indeed, the proof of the Far Reduction Theorem is almost the same as the proof of the theoretical part of the Descent Lemma II.

5.8 Discussion

Recall that \widehat{A} and \widehat{B} are partitioned into atoms such that the first return map of $\widehat{\Psi}$ is entirely defined on the interiors of these atoms. Experimental evidence leads us to the following conjecture.

Conjecture 5.7 *No point on the boundary of an atom has a well defined $\widehat{\Psi}$ orbit.*

Conjecture 5.7 is subtle, because there are boundary points on which at least 2^{18} iterates of Ψ , both forwards and backwards, are well-defined. In particular, thousands of iterates of the first return map $\widehat{\Psi}|_{\widehat{A}}$ are well-defined on such points.

Were the conjecture true, we could eliminate the restriction to generic orbits. The vertices of the atoms in the Renormalization Theorem all lie in $\mathbf{Z}[\phi]^3$. Given the formula for the map $\hat{\Theta}$, the preimage in Σ of these faces is a countable discrete set of non-horizontal lines defined over $\mathbf{Z}[\phi]$. For this reason, Θ never maps a point of a generic orbit into the boundary of one of the atoms. This is why we work with generic orbits.

6 Applications

6.1 Proof of Theorem 1.5

We will first consider the case of Theorem 1.5 in which $y_2 = R(y_1)$. Here R is the circle renormalization map. After we finish this case, we will deal with the general case. Note that the winding number of an orbit coincides with its Ψ -period. So, we will work with the Ψ -period rather than the winding number. We will often just say *period* in place of Ψ -period.

Suppose that O_1 and O_2 respectively are generic orbits having sufficiently high period. We have already explained the renormalization map $O_1 \rightsquigarrow O_2$. This map is 3-to-1 on the level of orbits, and the inverse images of a single orbit lie in different horizontal planes. So, in our case, we get a bijection between a certain collection of orbits on $\mathbf{R}_{y_1}^2$ and a certain collection of orbits in $\mathbf{R}_{y_2}^2$. These collections contain all generic infinite orbits and all generic periodic orbits of sufficiently high period.

Lemma 6.1 *O_1 and O_2 are coarsely equivalent, and the coarse equivalence constant is completely uniform – not even dependent on the parameters.*

Proof: We use the language of the Far Reduction Theorem. We decompose O_1 into \widehat{A} -cores $\{\alpha_k\}$. At the same time we decompose O_2 into \widehat{B} -cores $\{\beta_k\}$. We set things up so that $\alpha_k \rightsquigarrow \beta_k$ for all $k \in \mathbf{Z}$. Let λ be either 1 or ϕ^{-3} . According to the Far Reduction Theorem, the map $(x, y) \rightarrow (\lambda x, y)$ carries α_k to β_k , up to a uniformly bounded error, independent of y_1 and y_2 . Our result follows immediately from this. ♠

Remark: We have established the existence of a coarse equivalence constant that does not depend on the parameters, and this is even stronger than what we are claiming in Theorem 1.5. The difference is that the parameters in Theorem 1.5 might be related by a long chain of renormalizations, and the coarse equivalence constants will probably depend on the length of the chain. In view of the freezing phenomenon discussed in §4.5, there couldn't possibly be a uniform constant that worked for all equivalent parameters.

Now we turn to the question of local equivalence.

Lemma 6.2 *Corresponding orbits are locally equivalent.*

Proof: In the periodic case, each orbit is contained inside a periodic tile and there is nothing to prove. So, suppose that $O_1 \rightsquigarrow O_2$ and these orbits are both infinite. We think of the strip Σ as a subset of $\widehat{\Sigma}$. With this interpretation, the action of Ψ on Σ coincides with the action of $\widehat{\Psi}$ on $\widehat{\Sigma}$. So, in our proof we can think of O_1 and O_2 as orbits of $\widehat{\Psi}$. Every orbit is homogeneous, so it suffices to produce points $p_j \in O_j$, a disk Δ_j containing p_j , and a similarity from Δ_1 to Δ_2 which maps p_1 and p_2 and (generically) conjugates $\widehat{\Psi}|_{\Delta_1}$ to $\widehat{\Psi}|_{\Delta_2}$.

We can just choose $p_1 \in \widehat{A}$ and $p_2 \in \widehat{B}$ to be points are related by \widehat{R} . We choose Δ_1 small enough so that $\Delta_1 \subset \widehat{A}$, and then we let $\Delta_2 = \widehat{R}(\Delta_1)$. By the Renormalization Theorem, \widehat{R} conjugates $\widehat{\Psi}|_{\Delta_1}$ to $\widehat{\Psi}^{\pm 1}|_{\Delta_2}$ on generic points. This is exactly what we need. ♠

Now we turn to the general case of Theorem 1.5. Let $\mathcal{O}(y)$ denote the set of generic orbits in \mathbf{R}_y^2 . Say that a *full subset* of $\mathcal{O}(y)$ is a set that contains all infinite orbits and all periodic orbits having sufficiently high Ψ -period.

Reflection in the x -axis conjugates Ψ to Ψ^{-1} , and sets up an obvious and canonical bijection between $\mathcal{O}(y)$ and $\mathcal{O}(-y)$. (Here we think of $-y$ as an element of $\mathbf{R}/2/\mathbf{Z}$.) So, for Theorem 1.5, it suffices to consider the case of two parameters that equivalent under the group G_2^+ studied in §2.3.

By the Descent Lemma II, two such parameters y_1 and y_2 are such that

$$y = R^{n_1}(y_1) = R^{n_2}(y_2).$$

Let $y'_1 = R(y_1)$ and $y''_1 = R(y'_1)$, etc. The work above gives us a bijective correspondence between full subsets of $\mathcal{O}(y_1)$ and $\mathcal{O}(y'_1)$. Similarly, we get a bijection between $\mathcal{O}(y'_1)$ and $\mathcal{O}(y''_1)$. And so on. Composing all these bijections, we get a bijection between a full subset of $\mathcal{O}(y_1)$ and a full subset of $\mathcal{O}(y)$. Call this *the first main bijection*. From the lemmas proved above, corresponding orbits are locally and coarsely equivalent, and the coarse equivalence constant only depends on the parameter.

We get the same results for y_2 in place of y_1 . Call this the *second main bijection*. Composing the first and second main bijections, we get the bijection between full subsets of $\mathcal{O}(y_1)$ and $\mathcal{O}(y_2)$ which has all the properties advertised in Theorem 1.5. This completes the proof.

6.2 Bringing Orbits into View

Now we explore the consequences of the Near and Far Reduction Theorems. We use the same equivalence relation (\rightsquigarrow) as in the Near Reduction Theorem.

Lemma 6.3 *Let O_1 be any generic infinite Ψ orbit. There is a sequence of orbits $O_1 \rightsquigarrow \dots \rightsquigarrow O_n$ such that O_n intersects the rectangle $[-24, 24] \times [-2, 2]$. If O_1 intersects the region $[-N, N] \times [-2, 2]$ then we can take $n \leq 4N$.*

Proof: We define the type of an orbit to be the type of the \hat{A} -cores comprising it. Let O_1 be an infinite Ψ orbit. Let O_2 be such that $O_1 \rightsquigarrow O_2$.

We define $|S|_x$ exactly as in Equation 51, except that we use “inf” in place of “min” because we are dealing with an infinite set.

Suppose first that O_1 has type 2. and $|O_1|_x > 24$. Let α be any core and let β be such that $\alpha \rightsquigarrow \beta$. Far Reduction Theorem tells us that

$$|\beta|_x \leq \phi^{-3}|\alpha|_x + 15.$$

When $|\alpha|_x > 24$, a bit of arithmetic tells us that $|\beta|_x < |\alpha|_x - 1$. So, when $|O_1|_x > 24$, we have $|O_2|_x < |O_1|_x - 1$.

Suppose now that O_1 has type 1. In this case, given the nature of R , we have $O_1 \rightsquigarrow O_2 \rightsquigarrow O_3$, where O_2 has type 2. The Compression Theorem, combined with the same argument as above, tells us that

$$|O_2|_x \leq |O_1|_x + 12; \quad |O_3|_x = \phi^{-3}|O_2|_x + 15.$$

Therefore

$$|O_3|_x \leq \phi^{-3}|O_1|_x + 12\phi^{-3} + 15.$$

In this case, a bit of arithmetic tells us that $|O_3|_x < |O_1|_x - 1/2$ provided that $|O_1|_x > 24$. In either case, at most 2 renormalizations the closest point of O_1 closer by $1/2$ units. ♠

Combining Lemma 6.3 with the Near Reduction Theorem, we obtain the following result.

Theorem 6.4 *Let O_1 be any generic infinite orbit. Then there is some m such that $O_1 \rightarrow \dots \rightarrow O_m$, and O_m intersects the fundamental triangle T . More precisely, if O_1 is an unbounded orbit that intersects $[-2, 2] \times [-N, N]$ then we can take $n < 4N + C_y$. Here C_y is a constant that depends only on the value of $y \in \mathbf{R}/2\mathbf{Z}$ such that $O_1 \subset \mathbf{R}_y^2$.*

6.3 Proof of Theorem 1.4

Theorem 1.4 deals with the fundamental triangle T and the corresponding tiling \mathcal{T} . Recall from §3.4 that $\mathcal{T}^+ = \mathcal{T}$ and that \mathcal{T}^- is the image of \mathcal{T} under the reflection in the vertical line $x = 1$. We really only care about the tiling \mathcal{T} , but we find it convenient to consider both \mathcal{T}^+ and \mathcal{T}^- at the same time.

Lemma 6.5 *Let p be a generic point contained in the interior of a tile of \mathcal{T}^\pm . Then p has a periodic orbit. There is an upper bound on the period of p that depends only on the depth of the tile. Finally, the period of p tends to ∞ as the depth of the tile tends to ∞ .*

Proof: We use the language of the Fundamental Orbit Theorem. We check directly that the tiles of $\mathcal{T}(2)$ are periodic tiles. We handle the remaining tiles by induction.

Suppose that the first statement of the lemma holds for all tiles having depth at most $n - 1$. Let J_1 be a tile of depth n and let $p \in J_1$ be a generic point. There are indices i and j , and some choice of \pm , such that $J_1 \subset T_{ij}^\pm$ and $J' = R_{ij}^\pm(J)$ is a tile of depth $n - 1$. Let $J_2 = \psi^{-1}(J')$. Then the lemma holds for all generic points in J_2 . Let $s = \psi^{-1} \circ R_{ij}^\pm$ be the similarity carrying J_1 to J_2 . By the Fundamental Orbit Theorem, the orbit $p_1 \in J_1$ is periodic if and only if the orbit $p_2 = s(p_1) \in J_2$ is periodic. By induction p_2 is a periodic point. Hence, so is p_1 . This proves the first statement of the lemma.

For the second statement, observe that the process of renormalization shortens an orbit by at most a factor of 812. So, the period of generic point in a tile of depth n is at most 812^n . (This is a terrible estimate.)

For the third statement, we recall Statement 2 of the Renormalization Theorem: The renormalization operation shortens the orbit length. So, if p lies in a tile of depth n , then p has period at least n . (This is another terrible estimate.) ♠

Lemma 6.6 *Suppose $p \in T$ has a well-defined orbit but p does not lie in the interior of a tile of \mathcal{T} . Then p has an infinite orbit.*

Proof: If p had a periodic orbit then some neighborhood U of p would be such that all points in U were periodic, with the same period. But U necessarily contains infinitely many tiles of \mathcal{T} . This contradicts Lemma 6.5. ♠

Lemma 6.7 *No point on the boundary of tile of \mathcal{T} has a well-defined orbit.*

Proof: Lemma 6.6 shows that p cannot have a periodic orbit. On the other hand, suppose that $p \in \partial J$ has an infinite orbit, where J is a tile of \mathcal{T} . Then for any n , there is an open neighborhood U_n such that $p \in U_n$ and the first n iterates of Ψ are defined on p . These iterates all act by translation, so no point in U_n has period less than n . But U_n contains some generic points of J , no matter how large n . This contradicts the second statement of the previous lemma. ♠

Corollary 6.8 *Every tile of \mathcal{T} is a finite union of orbit tiles.*

Proof: Let J be a tile of \mathcal{T} . The same argument as in the first half of Lemma 6.6 shows that there is a uniform bound on the period of p . Let N be this bound. Since Ψ is a polygon exchange map, there is a *finite* set of lines in the plane such that the first N iterates of Ψ are defined in the complement of these lines. These lines partition J into finitely many smaller convex polygons, and Ψ is well defined and periodic on the complement of each of these lines. This shows that J is covered by a finite union of periodic tiles. Since no point of ∂J has a well-defined orbit, none of these periodic tiles crosses the boundary of J . ♠

Theorem 1.4 follows from the results above. We prove one more related result in this section.

Lemma 6.9 *Let L be any horizontal line segment such that $L \cap T^\pm$ contains no generic points with infinite orbits. Then there is a uniform upper bound to the period of any orbit on $L \cap T^\pm$.*

Proof: When the height of L lies in the interval I_1 , $L \cap T^\pm$ is contained in the closure of the union of the two largest octagonal tiles. The result is certainly true in this case. In general, we either have $R^n(y) = [0]$ or $R^n(y) \in I_1$ for some n . In either case, it follows from the symmetry of \mathcal{T} that $L \cap T^\pm$ is contained in the closure of the union of a finite number of orbit tiles. But then our bound follows from Lemma 6.5. ♠

6.4 Unboundedness

Theorems 1.5 and 6.4 are the workhorses in our overall proof. They allow us to transfer statements about orbits that intersect T , the fundamental triangle, to statements about orbits in general. The results in this section illustrate the technique.

We say that a *fundamental orbit* is an infinite orbit that intersects the fundamental triangle T .

Lemma 6.10 *Every fundamental orbit is unbounded in both directions.*

Proof: Every fundamental orbit is contained in the union of two lines, and every line intersects S , the fundamental fractal, in a nowhere dense set. Now we apply our unboundedness criterion, Lemma 4.4. ♠

Now we promote Lemma 6.10 to a statement about all orbits.

Lemma 6.11 *Every infinite orbit is unbounded in both directions.*

Suppose first that O is a generic infinite orbit. By Theorem 6.4, we can write $O = O_1 \rightarrow \dots \rightarrow O_n$, where O_n is an infinite orbit that intersects T and is coarsely equivalent to O_1 . But we have already seen that O_n is unbounded in both directions. Hence O_1 is also unbounded in both directions.

Now suppose that O is an infinite (non-generic) orbit that is bounded in, say, the forward direction. Then, by compactness, some point of O lies in the accumulation set of O . But then every point of O lies in the accumulation set of O . This implies that the set of accumulation points of O is uncountable. Indeed, this accumulation set contains a Cantor set.

The set of non-generic points on any horizontal line is countable, and O lies on 2 horizontal lines. (We take the Ψ -orbit, as usual.) Hence, there is some generic accumulation point p of O . Note that p necessarily has an infinite orbit. Hence p has an orbit that is unbounded in both directions.

In particular, for every N there is some ϵ such that any point within ϵ of p follows the orbit of p for N steps in either direction. Hence, p cannot be the accumulation point of an orbit that is bounded in one direction or the other. This contradiction shows that O is unbounded in both directions. ♠

6.5 Ruling out Some Heights

In this section we take care of a few annoying technical details that will make our arguments in the next section go more smoothly.

Lemma 6.12 *Let $y = \phi$. Then \mathbf{R}_y^2 has no generic unbounded orbits.*

Proof: By the Descent Lemma I (or a direct calculation), the point ϕ is 2-periodic with respect to R . We have $R(\phi) = 2 - \phi$. Let L_1 denote the horizontal line $y = \phi$. Let L_2 denote the horizontal line $y = 2 - \phi$. Let T be the fundamental triangle. Note that $L_1 \cap T$ is just a single point, and this point is non-generic. At the same time, $L_2 \cap T$ is contained entirely in the periodic tile K_0 . Hence neither line intersects T in any point that could have a generic unbounded orbit.

Suppose, on the other hand, that \mathbf{R}_y^2 has a generic unbounded orbit O_1 . By Theorem 6.4 there is some m such that $O_1 \rightarrow \dots \rightarrow O_m$ and O_m intersects T . The intersection $O_m \cap T$ is a generic unbounded orbit by Theorem 1.5. The operation of replacing an orbit by an associate simply switches from R_y^2 to R_{2-y}^2 . Hence, $O_m \cap T$ is a point of either $L_1 \cap T$ or $L_2 \cap T$. This is a contradiction. ♠

Corollary 6.13 *Suppose that $y \in C - C^\#$. Then \mathbf{R}_y^2 has no generic unbounded orbits.*

Proof: We have $y \in \mathbf{Z}[\phi]$ by the construction of C . By Lemma 2.7, we see that $y = m + n\phi$, with m even. But then $y \sim 0$ or $y \sim \phi$. Here \sim denotes G_2 -equivalence. The map Ψ is the identity on \mathbf{R}_0^2 . So, Theorem 1.5 rules out the possibility that $y \sim 0$. On the other hand, Theorem 1.5 and the preceding lemma rule out the possibility that $y \sim \phi$.

Corollary 6.14 *Any generic unbounded orbit that intersects the fundamental triangle T contains a point of the form (x, y) where $y \in C^\#$.*

Proof: By Theorem 1.4, and the structure of the fundamental fractal S , any such point (x, y) must have $y \in C$. But our previous result rules out the possibility that $y \in C - C^\#$.

6.6 Proof of Theorem 1.6

Now we turn to the question of self-accumulation. Lemma 4.3 tells us, in particular, that there is a canonical bijection between the Ψ orbits in Σ_+ and the Ψ orbits in Σ_- . We call two such orbits *partners*. Just for this section, we introduce the notation $O_1 \Rightarrow O_2$ to mean that $O_1 \rightsquigarrow O'_2$, where O'_2 is the partner of O_2 . The following result is just a reformulation of the Fundamental Orbit Theorem.

Corollary 6.15 *Let $p_1 \in T_{ij}^\pm - \mathcal{T}(2)$ be a generic point and let $p_2 = R_{ij}^\pm(p_1)$. Let O_1 and O_2 respectively be the orbits of p_1 and p_2 . Then $O_1 \Rightarrow O_2$.*

Proof: If O'_2 is an orbit in Σ_- , and $\psi(O'_2)$ intersects Σ_+ , then the partner of O_2 is the orbit of any point of $\psi(O'_2)$. The corollary follows from this fact, and from the Fundamental Orbit Theorem. ♠

Let U_y denote the set of unbounded orbits in \mathbf{R}_y^2 . We also have the following lemma, which is just a reformulation of part of Theorem 1.5.

Lemma 6.16 (Bijection Principle) *Let O_1 and O'_1 be two generic orbits in $U_y \cap \Sigma_+$ such that $O_1 \Rightarrow O_2$ and $O'_1 \Rightarrow O_2$. Then $O_1 = O'_1$.*

Proof: We have $O'_1 \rightsquigarrow O'_2$ and $O_1 \rightsquigarrow O'_2$ where O'_2 is the partner of O_2 . On the level of infinite orbits, the renormalization map is 3-to-1. At the same time, the 3 preimages of an orbit lie in different sets U_y . Since O_1 and O'_1 lie in the same set U_y , we have $O_1 = O'_1$. ♠

Lemma 6.17 *Let $p \in T$ be a generic fundamental orbit. Then the orbit of p is self-accumulating.*

Proof: We already know that the orbit of p is unbounded in both directions. By Lemma 6.14, we have $p = (x, y)$, where $y \in C^\#$. Let $O(p)$ denote the orbit of p . We refer to the notation and terminology in §3.5. Let $\Upsilon(p)$ denote the renormalization set for p . Let $q \in \Upsilon(p)$. Here p and q lie on the same horizontal line, and

$$p = p_0 \rightarrow p_1 \dots \rightarrow p_n; \quad q = q_0 \rightarrow q_1 \dots \rightarrow q_n; \quad p_n = q_n. \quad (52)$$

By the Corollary above,

$$O(p_0) \Rightarrow \dots \Rightarrow O(p_n); \quad O(q_{n-1}) \Rightarrow O(q_n) = O(p_n). \quad (53)$$

Since p_{n-1} and q_{n-1} lie on the same horizontal line, the Bijection Principle says that $O(q_{n-1}) = O(p_{n-1})$. It now follows from induction on n that $O(p) = O(q)$.

Since q is an arbitrary point of $\Upsilon(p)$ we now know that

$$\Upsilon(p) \subset O(p). \quad (54)$$

By the Density Lemma, $\Upsilon(p)$ is dense in $\Lambda = S \cap L$, where L is the horizontal line containing p . By the Horizontal Lemma, Λ is a Cantor set. Hence p is an accumulation point of $O(p)$. But, $O(p)$ is homogeneous. Hence, every point of $O(p)$ is an accumulation point of $O(p)$. This proves that $O(p)$ is self-accumulating. ♠

Now we prove Theorem 1.6. Let O_1 be a generic infinite orbit. By Theorem 6.4, we have $O_1 \rightarrow \dots \rightarrow O_n$ where O_n is an infinite orbit that intersects the fundamental triangle. But we have already shown that O_n is self-accumulating. By Theorem 1.5, the orbits O_1 and O_n are locally similar. Hence O_1 is also self-accumulating.

6.7 Proof of Theorem 1.1

We've already proved that every orbit is either periodic or unbounded in both directions. It only remains to show that the union U of the unbounded orbits has Hausdorff dimension 1. By Lemma 3.1 the set of generic points in S has Hausdorff dimension 1. All such points have well-defined orbits and, by Theorem 1.4, all the orbits are unbounded. Hence U contains a 1 dimension set. Hence $\dim(U) \geq 1$.

Let G be the set of generic points in \mathbf{R}^2 . note that $\mathbf{R}^2 - G$ has Hausdorff dimension 1 because it is a countable set of lines. Combining Theorems 1.5 and 6.4, we see that every generic point $p \in U$ has a neighborhood Δ such that $\Delta \cap U$ is similar to a subset of S . Hence $\dim(U \cap G \cap \Delta) \leq 1$. But we can cover any compact subset of U by finitely many such neighborhoods. Hence $\dim(U \cap G) \leq 1$. Since $U - G$ is contained in a countable family of lines, we have $\dim(U - G) \leq 1$. Hence $\dim(U) \leq 1$.

6.8 Proof of Theorem 1.2

First we establish part of Theorem 1.9. Let U_y^* denote the set of generic unbounded orbits in \mathbf{R}_y^2 .

Lemma 6.18 *U_y is empty if U_y^* is empty.*

Proof: Let R be the renormalization map. Suppose $R^n(y) = 0$ for some n . All orbits on the line $y = 0$ have period 1, and renormalization decreases periods by at most a factor of 812. Hence, there is a uniform bound to the period of any generic point in U_y .

On the other hand, suppose that y is not in the inverse image of 0. Let p_1 be any generic point that is, say, within 1 unit of y . Let O_1 be the orbit containing p_1 . Combining Lemma 6.3 and the second statement of the Near Reduction Theorem, we see that there is some uniform m such that $O_1 \rightarrow \dots \rightarrow O_m$ and O_m intersects T . But the horizontal line containing $O_m \cap T$ has no generic points with infinite orbits. Hence, by Lemma 6.9, there is a uniform upper bound to the period of O_m . Hence, there is a uniform upper bound to the period of O_1 .

Finally, we can take a sequence of generic points approximating a supposed infinite orbit. This is incompatible with the uniform upper bound we have on the periods of this approximating sequence. ♠

Now we prove Theorem 1.2. Let $Y \subset \mathbf{R}/2\mathbf{Z}$ denote those y such that \mathbf{R}_y^2 contains unbounded orbits. Let $y \in C^\#$. The horizontal line of height y intersects $S^\#$ in an uncountable set which must have points with well-defined orbits. By Theorem 1.4, these points are not periodic. By Lemma 6.10, these points have unbounded orbits. Hence $C^\# \subset Y$. Theorem 1.5 now shows that Y contains all points y such that $y \sim c$ and $c \in C^\#$.

Now for the converse. Suppose that $y_1 \in Y$. By Lemma 6.18, we can assume that $\mathbf{R}_{y_1}^2$ has a generic unbounded orbit O_1 . By Theorem 6.4, we have $O_1 \rightarrow \dots \rightarrow O_m$ with O_m intersecting T . Let y_k be such that O_k is an orbit of $\mathbf{R}_{y_k}^2$. Passing to associates preserves the G_2 -equivalence class, and so does R . Hence $y_1 \sim y_m$.

By Theorem 1.5, the orbit O_m is generic and unbounded. By Theorem 1.4, we have $y_m \in C$. By Lemma 6.14, we have $y_m \notin C - C^\#$. Hence $y_m \in C^\#$ and $y_1 \sim y_m$, as desired. Finally, we compute easily that $\dim(C) = \log(3)/\log(\phi^3)$.

6.9 Proof of Theorem 1.3

As we remarked above, the concepts of winding number and Ψ -period coincide. We will work with Ψ -period, as usual.

Lemma 6.19 *Every horizontal line in Σ contains a dense set of periodic orbits.*

Proof: Let U denote the union of unbounded orbits. If $L \cap U$ is empty, there is nothing to prove. Otherwise, by Theorem 6.4 and Theorem 1.5, the set $L \cap U$ is locally similar to $L' \cap U$, where L' is a line segment in T , the fundamental triangle. But $U \cap T$ is nowhere dense in each horizontal line. Hence $U \cap L'$ is nowhere dense in L' . Hence $U \cap L$ is nowhere dense in L . ♠

Lemma 6.20 *Let $y \in \mathbf{R}/2\mathbf{Z}$ be any value other than*

$$0; \quad 4 - 2\phi; \quad -2 + 2\phi; \quad 2$$

Then \mathbf{R}_y^2 contains a periodic orbit that intersects \widehat{B} , the renormalization set.

Proof: Let $y \in \mathbf{R}/2\mathbf{Z}$ be any value other than the ones listed. Let Π be the horizontal plane of height y . The plane Π intersects \widehat{B} in an open set. Let $L \subset \Sigma$ be the horizontal line of height y . Since $\Theta(L)$ is dense in Π , there is an open subset $V \subset L$ such that $\Theta(V) \subset \widehat{B}$. By Lemma 6.19, there is a dense subset of V consisting of periodic points. Any periodic point in V works for us. ♠

Lemma 6.21 *Let $y \in \mathbf{R}/2\mathbf{Z}$. Suppose it never happens that $R^n(y) = [0]$. Then \mathbf{R}_y^2 contains periodic points having arbitrarily high Ψ -period.*

Proof: Our proof refers to the terminology used in the Renormalization Theorem. Let $y_n = R^n(y)$. For our analysis, we identify Σ with a dense subset of $\widehat{\Sigma}$. Define $\Sigma_n = \mathbf{R}_{y_n}^2 \cap \Sigma$. Let $\widehat{\Sigma}_n$ be the horizontal plane of height y_n .

Referring to Lemma 6.20, the heights of the excluded horizontal planes all lie in the inverse image of 0. For this reason, $\widehat{\Sigma}_n$ contains a periodic tile P whose orbit intersects \widehat{B} in an atom that is a translate of P .

Since Σ_n is dense in $\widehat{\Sigma}_n$, the periodic tile P must intersect Σ_n in an open interval. In particular Σ_n contains a generic periodic point whose orbit O_n intersects \widehat{B} . By the Renormalization Theorem, we can find a generic orbit O'_{n-1} in Σ_{n-1} such that $O'_{n-1} \rightsquigarrow O_n$. The period of O'_{n-1} is longer than the period of O_n .

O'_{n-1} necessarily intersects \widehat{A} . By the Renormalization Theorem, O'_{n-1} also intersects \widehat{B} . So, we can set $O_{n-1} = O'_{n-1}$ and repeat the above argument to produce a generic periodic orbit O'_{n-2} in Σ_{n-2} that intersects \widehat{B} . And so on. The final orbit O_1 has period at least $n-1$. But n is arbitrary. ♠

The full inverse image of $[0]$ is precisely $2\mathbf{Z}[\phi]$. Hence, if $y \notin 2\mathbf{Z}[\phi]$, then \mathbf{R}_y^2 contains orbits of arbitrarily high Ψ -period.

Lemma 6.22 *Suppose that $y = m + n\phi$ where both m and n are even. Then there is a uniform bound on the Ψ -period of any point of \mathbf{R}_y^2 with a well defined orbit.*

Proof: We consider generic points first. Suppose \mathbf{R}_y^2 contains a generic orbit with an enormous winding number – either finite or infinite. There is some n such that $R^n(y) = [0]$. If the winding number of our orbit is too large, then we can renormalize this orbit more than n times. But this contradicts the fact that Ψ is the identity on $\widehat{\Sigma}_0$, and no such orbit on this plane has a renormalization.

So, every generic point of \mathbf{R}_y^2 is periodic and we have a uniform bound on the periods. But any point in \mathbf{R}_y^2 can be approximated by a sequence of generic points. The uniform bound on the approximating sequence immediately gives the same uniform bound on the limit, provided that the limit has a well-defined orbit. ♠

Let $y \in 2\mathbf{Z}[\phi]$ be some point for which the following property $(*)$ holds:

$$R^k(y) \neq [1]; \quad k = 0, \dots, n. \quad (55)$$

Then the same argument as in Lemma 6.21 shows that \mathbf{R}_y^2 contains generic points having Ψ -period at least $n-1$. Finally, for any fixed n , there are only finitely many points of $2\mathbf{Z}[\phi]$ which fail to satisfy Equation 55. This establishes Theorem 1.3.

6.10 Proof of Theorem 1.10

This is one of the more subtle results. We will try to break the proof down into small steps.

Lemma 6.23 *Let y be one of the 4 values*

$$-1 + \phi \quad -3 + 3\phi, \quad 3 - \phi, \quad 5 - 3\phi.$$

Let Λ be the intersection of the line $\mathbf{R} \times \{y\}$ with the fundamental triangle T . Then there is a uniform upper bound to the period of any point that is sufficiently close to Λ .

Proof: Let $\Lambda_1, \dots, \Lambda_4$ be the 4 line segments in question. We check by direct inspection that there is some $\epsilon_0 > 0$ with the following property. The ϵ_0 neighborhood of Λ_k is contained in a finite union of tiles of \mathcal{T} . Indeed, the two lines Λ_2 and Λ_4 are disjoint from T , the interior of $\Lambda_1 \cap T$ is contained in the interior of a single periodic tile, and the interior of $\Lambda_3 \cap T$ is contained in the interior of the closure of two periodic tiles. ♠

Let $y = m + n\phi$ where m and n are odd. Let $L = \mathbf{R} \times \{y\}$. Let $N = N_y$ be as in Lemma 2.5.

Lemma 6.24 *Let $k > N$ be a fixed integer. There is some $\epsilon > 0$ with the following property. Suppose that O_1 is an orbit that contains a point within ϵ of L . Suppose $O_1 \rightarrow \dots \rightarrow O_k$ and that O_k intersects the fundamental triangle. Then there is a uniform bound on the period of O_1 .*

Proof: By Lemma 2.5, and the fact that R^k expands distances by at most ϕ^{3k} , the orbit O_k must intersect the fundamental triangle in a point very close to one of the 4 line segments listed in Lemma 6.23, provided that ϵ is small enough. The point here is that, when we keep track of which set \mathbf{R}_y^2 contains our successive orbits, the renormalization operation implements the map R and switching from an orbit to an associate implements the map S .

If O_k is very close to one of the segments in Lemma 6.23, then there is a uniform upper bound to the period of O_k . But the renormalization operation, and the operation of switching to an associate, only can decrease the period by a uniformly bounded factor. Hence, there is a uniform upper bound on the period of O_1 . ♠

Lemma 6.25 *Let $p \in L$ be some point. Let M be any positive integer. Then there is some $\epsilon > 0$ with the following property. If O_1 is a generic periodic orbit that comes within ϵ of p , then there is some $k > M$ such that $O_1 \rightarrow \dots \rightarrow O_k$ and O_k intersects the fundamental triangle. The same k works for all orbits satisfying the hypotheses.*

Proof: Combining Lemma 6.3 and the second statement of the Near Reduction Theorem, we see that there is some k , which works for all choices of O_1 , such that $O_1 \rightarrow \dots \rightarrow O_k$ and O_k intersects the fundamental triangle. The only trouble is that we might have $k < M$. However, we can renormalize again and then, if necessary, apply Theorem 6.4 again. This produces a larger value of k . We keep going like this until we arrive at some $k > M$, and then we stop. ♠

Corollary 6.26 *Let $p \in L$ be any point. Then there is some $\epsilon > 0$ and some constant Z with the following property. If O_1 is a generic periodic orbit that comes within ϵ of p then O_1 has period at most Z .*

Proof: Choose $M = N$, the constant in Lemma 6.24. If ϵ is small enough then we have some fixed $k > M$ such that $O_1 \rightarrow \dots \rightarrow O_k$ and O_k intersects the fundamental triangle. This fixed k works for any orbit that comes within ϵ of p . But now Lemma 6.23 applies to O_1 . ♠

Suppose that the conclusion of Theorem 1.10 is false. Then we can find some $p \in L$ which is the accumulation point of points having unbounded orbits. By Theorem 1.1, we also have a sequence of generic periodic points converging to a point of L whose period tends to ∞ . This contradicts the corollary we have just proved. This establishes This proves Theorem 1.10.

We mention the following corollary of Theorem 1.10.

Corollary 6.27 *Let $y = m + n\phi$ with m and n odd. Let K be and compact subset of \mathbf{R}^2 . Then there is some $\epsilon > 0$, depending on K, n, m , such that no point of K within ϵ of the line L_y has an unbounded orbit.*

Proof: This is just an application of compactness. ♠

6.11 Proof of Theorem 1.8

Let Y_r be as in Theorem 1.8. As above, let U_y denote the union of unbounded orbits in \mathbf{R}_y^2 .

Lemma 6.28 *Y_r is nowhere dense.*

Proof: Rather than work precisely with the set mentioned in Theorem 1.8, we define Y'_r to be the set $y \in Y$ such that U_y contains a generic orbit that intersects the rectangle $[-r, r] \times [-2, 2]$. We have $Y_{r-1} \subset Y'_r \subset Y_{r+1}$, so it suffices to prove Theorem 1.8 for Y'_r in place of Y_r .

The analysis in Lemma 6.3 shows that

$$R^n(Y'_r) \cup R^{n+1}(Y'_r) \subset Y'_{24}; \quad \forall n > 4r. \quad (56)$$

Suppose that Y'_r is not nowhere dense for some r . Since R^2 is an expanding map, and Y'_r is not nowhere dense, the left hand side of Equation 56 is dense in $\mathbf{R}/2\mathbf{Z}$ for sufficiently large n . Hence Y'_{24} is dense in $\mathbf{R}/2\mathbf{Z}$. But this contradicts Corollary 6.27. ♠

We have already computed that $\dim(C^\#) = \log(3)/\log(\phi^3)$. Observe that every neighborhood of the Penrose kite vertex $(\phi^{-3}, 0)$ contains a similar copy of S . Hence $\dim(Y_r) = \dim(C^\#)$.

6.12 Proof of Theorem 1.9

When we proved Theorem 1.2, we proved the first statement of Theorem 1.9. Now we prove the second.

The set of non-generic points is contained in a countable union of lines. In particular, the set of non-generic unbounded orbits is contained in a countable union of sets of the form

$$L \cap (\mathbf{R} \times Y), \quad (57)$$

where Y is the set of y such that \mathbf{R}_y^2 has unbounded orbits. By Theorem 1.2, we have $\dim(Y) = \log(3)/\log(\phi^3)$. But then the union of non-generic unbounded orbits is contained in a countable union of sets that have the same dimension as Y . This completes the proof.

6.13 Proof of Theorem 1.7

Almost every statement in Theorem 1.7 is an immediate consequence of Theorems 1.4, 1.5, and 6.4. The one statement that is not immediate is that a neighborhood of the dynamical tiling about a generic point with unbounded orbit is *isometric* to a small patch of \mathcal{T} .

The results mentioned above only imply that the abovementioned neighborhood is *similar* to a patch of \mathcal{T} . However, tracing through our argument, we see that the similarity factor is ϕ^{3k} for some integer k . But \mathcal{T} is a self-similar set with expansion constant ϕ^3 . So, we can take $k = 0$ in the similarity factor between the neighborhood of interest to us and a suitable patch of \mathcal{T} .

6.14 Proof of Theorem 1.11

Let $y = m + n\phi$. We already know that $1 \in C^\#$. So, by Theorem 1.2, we know that \mathbf{R}_y^2 has unbounded orbits provided that m is odd and n is even.

Conversely, suppose that \mathbf{R}_y^2 has unbounded orbits. The set $C - C^\#$ contains points of the form $m + n\phi$ with m even and n having either parity. Combining Theorem 1.2 and Corollary 6.14, we see that m cannot be even. When m is odd and n is odd, Theorem 1.10 shows that \mathbf{R}_y^2 has no unbounded orbits in this case. The only case left is when m is odd and n is even, as claimed.

7 Computational Methods

7.1 Golden Arithmetic

A large percentage of our proofs involve exact computer calculations, done over the number ring $\mathbf{Z}[\phi]$. We abbreviate these kinds of computer calculations as *golden arithmetic*. Here we describe how the computer does golden arithmetic, for the sake of making our calculations completely reproducible. There exist computer packages, such as Pari, which perform these kinds of calculations in extreme generality. A software package like Pari is vastly more complicated than our own much more limited collection of routines.

We represent the number $a_0 + a_1\phi$ as the integer pair (a_0, a_1) . The computer can only represent finitely many such numbers, but the large finite set of representable numbers is suitable for our purposes.

The Ring Operations: We have the following obvious rules

$$\begin{aligned}(a_0, a_1) \pm (b_0, b_1) &= (a_0 \pm b_0, a_1 \pm b_1); \\ (a_0, a_1) \times (b_0, b_1) &= (a_0b_0 + a_1b_1, a_0b_1 + a_1b_0 + a_1b_1),\end{aligned}\tag{58}$$

which represent the ordinary ring operations in $\mathbf{Z}[\phi]$. As long as all integers stay less than, say, 10^6 in absolute value, the computer adds and multiplies them correctly. The Galois map $\tau(a + b\phi) = a - b/\phi$ is a ring automorphism of $\mathbf{Z}[\phi]$. In terms of our representation, we have

$$\tau : (a, b) \rightarrow (a + b, -b).\tag{59}$$

Division: Since $\mathbf{Z}[\phi]$ is not a field, we cannot generally perform division in $\mathbf{Z}[\phi]$. However, it occasionally happens that we know in advance that $\gamma = \alpha/\beta$ lies in $\mathbf{Z}[\phi]$ and we want to find γ given α and β . In this cases, we find γ by computing

$$\gamma = \frac{\alpha \times \tau(\beta)}{\beta \times \tau(\beta)}.\tag{60}$$

The number in the denominator is an integer, and we find the quotient on the right hand side by dividing the coefficients of the numerator by this integer. Before returning the value, we verify that the resulting element belongs to $\mathbf{Z}[\phi]$ and satisfies the equation $\alpha = \beta \times \gamma$.

Positivity: Now we explain how we check that $a_0 + a_1\phi$ is positive. Our method, applied to $a_0 + a_1\phi$, returns a value of *true* only when $a_0 + a_1\phi > 0$. We call $a_0 + a_1\phi$ *strongly positive* if

$$a_0f_{100} + a_1f_{101} > 0; \quad a_0f_{101} + a_1f_{102} > 0. \quad (61)$$

Here f_n is the n th Fibonacci number. Since the successive quotients of Fibonacci alternately over and under approximate ϕ , the fact that both linear combinations are positive guarantees that $a_0 + a_1\phi$ is also positive. In short, a strongly positive element of $\mathbf{Z}[\phi]$ is positive. Certainly there are positive numbers that are not strongly positive. However, we do not encounter these numbers in our calculations.

In Equation 61, we use the BigInteger class in Java, which does exact integer arithmetic for integers up to many thousands of digits long. We might have used the BigInteger class for all our calculations, but this would make the calculations much slower and the computer routines much more tedious to program. Thus, we use the BigInteger class only when needed.

Equation 61 is the basis for all our computations that involve inequalities between elements of $\mathbf{Z}[\phi]$. For instance, to verify that (a_0, a_1) represents a number larger than (b_0, b_1) , we apply the test to (c_0, c_1) , where $c_j = a_j - b_j$.

Golden Structures: The *GoldenReal*, a pair (a_0, a_1) as above, is our basic object. We also define more complicated objects based on the GoldenReal:

- A *GoldenComplex* is a pair $x + iy$, where x and y are both GoldenReals.
- A *GoldenVector* is a tuple of GoldenReals.
- A *GoldenPolytope* is a finite list of GoldenVectors.

The usual operations for these objects are done using the ring operations described above.

Remark: It is worth mentioning that our graphical user interface mainly operates with floating point arithmetic, for the purposes of speed. We mainly use the special arithmetic when we need to do rigorous calculations for the purpose of making a proof. The special structures are all isolated in separate files, so as not to interfere with the rest of the program.

7.2 Overflow Error

There is one computational issue that we must face when we do exact integer arithmetic calculations. The computer does not reliably perform the arithmetic operations on very large integers. For example, when multiplying together ordinary integers, or *ints*, our computer tells us that

$$100000 \times 100000 = 1410065408.$$

Each *int* is a length 32 binary string. One of the bits records the sign of the integer and the remaining bits give the binary expansion. The problem with the solution to 100000×100000 is that it requires more than 32 bits to express the answer.

To give us a bit more flexibility, we use *longs*. A *long* is a 64-bit representation of an integer. The calculation above comes out right when we use *longs* in place of *ints*.

Here are some conservative bounds on what the computer can do with *longs*

- The computer reliably computes the value $a_1 \pm a_2$ provided that we have $\max(|a_1|, |a_2|) < 2^{60}$.
- The computer reliably computes the value $a_1 \times a_2$ provided that we have $\max(|a_1|, |a_2|) < 2^{60}$.

Whenever we perform one of the basic operations above, we check that the conditions above hold. The code is set up to interrupt the calculation if one of the conditions above is not met.

The only other operations we perform are Galois conjugation within $\mathbf{Z}[\phi]$ and the guided form of division mentioned above. For conjugation, we make sure that the coefficients of the GoldenReal are less than 2^{60} in absolute value.

The division operation requires a special explanation. When we perform the division operation to *find* the quotient $c = a/b$, we do not bound the sizes of the coefficients. Rather, when we then verify that indeed $a = b \times c$, we check the bounds of the coefficients of b and c . This means that the equality $a = b \times c$ has really been verified.

7.3 Some Basic Operations and Tests

Here we explain some basic operations we perform on polygons and polyhedra. When we have golden polygons and golden polyhedra, the operations are all done using the golden arithmetic described above.

Positive Convexity: We say that a polygon P is *positively convex* if P is convex, and if the orientation of ∂P given by the ordering on the vertices is counterclockwise. P is positively convex if and only if

$$\operatorname{Im}(\overline{z_{21}}z_{31}) \geq 0; \quad z_{ij} = z_i - z_j. \quad (62)$$

for every triple z_1, z_2, z_3 of consecutive vertices of P . Note that a positively convex polygon need not be strictly convex.

Containment Test for Polygons: Let P be a positively convex polygon with vertices z_1, \dots, z_k . A point w lies in the interior of P if and only if

$$\operatorname{Im}\left(\overline{(z_i - w)}(z_{i+1} - w)\right) > 0; \quad (63)$$

holds for every index i , with indices taken mod k .

Strict Convexity for Polyhedra: To check that a golden polyhedron is strictly convex, we exhibit a golden vector $V = V(P, v)$ such that $v \cdot P > v' \cdot V$ for all other vertices v' of P . In practice, we search for V amongst all golden vectors whose coefficients have real and imaginary parts of the form $a + b\phi$ with $\max(|a|, |b|) \leq 2$. In general, one would need to make a more extensive search or else have some special information about P .

Golden Interior Points: The most natural interior point of a polytope is its center of mass. However, the center of mass of a golden polytope might not be a golden vector. Here we describe a less canonical way of picking an interior point of a golden polytope which results in a golden vector.

Given two golden vectors V_1 and V_2 , the vector $a(V_1, V_2) = \phi^{-1}V_2 + \phi^{-2}V_1$ is a golden vector that lies on the line segment $\overline{V_1V_2}$. Note generally, suppose we have a golden polytope with vertices V_1, \dots, V_n . We define

$$W_2 = a(V_1, V_2); \quad W_3 = a(W_1, V_3); \quad \dots \quad W_n = a(W_1, V_n). \quad (64)$$

then W_n is a golden vector in the interior of the polytope.

Disjoint Interiors I: Here is one method we use to verify that two golden polyhedra P and P' have disjoint interiors. We exhibit a golden vector W such that $V \cdot W < V' \cdot W$ for all pairs (V, V') , where V is a vertex of P and V' is a vertex of P' . In practice, we search for W amongst all vectors $E \times E'$, where E is an edge of P and E' is an edge of P' .

Disjoint Interiors II A second way we will know (without directly testing) that two distinct polygons or polyhedra P_1 and P_2 have disjoint interiors is that there is some map $f : P_k \rightarrow \mathbf{R}^2$ such that f is entirely defined and affine on the interior of P_k , but f does not extend to be defined and affine in a neighborhood of any point of ∂P_k . If P_1 and P_2 did not have disjoint interiors, then some point of ∂P_1 lies in the interior of P_1 , and then the map f is then defined and locally affine in a neighborhood of this boundary point. This is a contradiction.

Coplanarity Test: Let V_1, \dots, V_k be finite list of golden vectors. Consider the successive vectors

$$N_i = (V_{i+1} - V_i) \times (V_{i+2} - V_i), \quad (65)$$

with indices taken mod k . The vectors are coplanar if and only if N_i and N_{i+1} are parallel for all i .

Raw Face Enumeration: Here is how we enumerate the faces of a polyhedron that is given in terms of its vertices. Let $V(P)$ be the vertex set of a polyhedron P . For each subset $S \subset V(P)$, having at least 3 elements, we first check if S is a coplanar set of vertices, as explained above. If S is not coplanar, we eliminate S from consideration.

Assuming that S is coplanar, let V_1, \dots, V_k be the vertices of S . Let N_1 be the vector from Equation 65. Let $V = V_1 + \dots + V_k$. Also, let kS denote the result of scaling all vectors of S by a factor of K . Likewise define kP . We try to check that either $V \cdot N_1 \leq W \cdot N_1$ for all vertices W of kP or that $V \cdot N_1 \geq W \cdot N_1$ for all vertices W of P . If one of these two things is true, then the center of mass of the convex hull of S is contained in ∂P , and this means that S is a subset of a face of P . The reason why we scale everything by k is that we want to work entirely in $\mathbf{Z}[\phi]$. (To avoid scaling, we could have used the golden interior point described above in place of V , but we didn't.)

Edge Enumeration: Given the list of vertices and the raw list of faces, we find the edges as follows. A pair of distinct vertices (v_1, v_2) is an edge of P if and only if P has 4 distinct vertices v_1, v_2, w_1, w_2 such that (v_1, v_2, w_1) and (v_1, v_2, w_2) are two distinct faces of P . For each pair (v_1, v_2) , we try to find the pair (w_1, w_2) with the property just mentioned. If we succeed, we add (v_1, v_2) to the list of edges.

Polished Face List: There is one more step, which we perform for the sake of tidiness, and also for the sake of making some of our other computations go more quickly. We would like the vertices of each face to be listed so that they go in cyclic order around the face. Our convention is that the vectors produced by Equation 65 should be outward normals. To order the vertices around a face, we first use the raw face list to enumerate the edges of the face. We then use the edges, in a fairly obvious way, as a guide for placing the vertices.

Slicing: As an application of our edge and face enumeration, we explain how we compute the intersection of a golden polyhedron P and a horizontal plane Π that contains no vertices of P . Every edge of P will satisfy Equation 33 and Π will have height in $\mathbf{Z}[\phi]$, so $P \cap \Pi$ will be a golden polygon.

Every edge of P either crosses Π at an interior point or else is disjoint from Π . We enumerate those edges e_1, \dots, e_k of P which cross Π . We order the edges so that e_i and e_{i+1} lie in a common face. Finally, we set $v_i = \Pi \cap e_i$. Then v_1, \dots, v_k are the vertices of the polygon $P \cap \Pi$, and they are cyclically ordered around the boundary of $P \cap \Pi$.

Containment Tests for Polyhedra: Let V be a point and let P be a polyhedron. Let F_1, \dots, F_k be the faces of P . Let W_1, \dots, W_k be such that W_j is a vertex of F_j for each j . Finally, let N_j be a vector that is perpendicular to F_j . We compute N_j by applying Equation 65 to the first 3 vertices of F_j . The point V lies in the interior of P if and only if $V \cdot N_j < V \cdot W_j$ for $j = 1, \dots, k$. Similarly, V lies in ∂P if and only if $V \cdot N_j \leq V \cdot W_j$ for all j , and there is equality for at least one index j .

Remark: The tests we have described above look fairly intensive, but for the most part we will be applying them over and over to the same list of 64 polyhedra – the ones that define our polyhedron exchange map. For these, we pre-compute everything and store the results in a look-up table.

7.4 Covering Tests

Here we describe how we show that a given polygon or polyhedron is covered by a finite union of other polygons or polyhedra. We will explain the polygon case first. We then reduce the polyhedron case to the polygon case by a trick involving slicing.

Let Q be a golden polygon and let P_1, \dots, P_n be a finite list of golden polygons. Here we explain the computational test we perform in order to prove that $Q \subset \bigcup P_k$.

Eliminating Spanning Edges: Say that P_k has a *spanning edge* relative to Q if some edge of P_k intersects Q in an interior point, but has no vertices in the interior of Q . In practice, we eliminate all spanning edges by adding an extra “vertex” of P_k to each edge. Since we want our polygons to remain golden, we pick a golden interior point for each edge. See above.

Our method is not guaranteed to eliminate all spanning edges, but in practice it does. In all cases but one, we also have $P_k \subset Q$ for all k . In this case, the only spanning edges are those that have both vertices on ∂Q . For such spanning edges, the addition of a single vertex does the job. In the one remaining case, namely the one involving the rectangle $Q = \Sigma^*$ in the next chapter, there are no spanning edges at all.

The Matching Property: Let v be an interior vertex of some P_i . Let P_1, \dots, P_k be the (re-indexed) union of all the polygons P_i such that $v \in \partial P_i$. When v is a vertex of P_i , we let e_i^+ be the edge of P_i such that v is the leading vertex, and we let e_i^- be the edge of P_i such that v is the trailing vertex. When p lies in the interior of an edge e_i of P_i , we let e_i^+ and e_i^- be the two oriented copies of e_i , labelled so that the indices of the endpoints of e_i^+ are cyclically increasing. We think of these $2k$ edges as vectors, all oriented away from v . For each $i \in \{1, \dots, k\}$ we check computationally that there is exactly one $j \in \{1, \dots, k\}$ such that the two vectors e_i^- and e_j^+ are parallel. Call this *the matching property*.

Lemma 7.1 *Suppose P_1, \dots, P_k is a collection of positively convex polygons having pairwise disjoint interiors. Suppose also that each P_k intersects Q and has no spanning edges with respect to Q . Finally, suppose that the matching property holds for each vertex of P_k that lies in the interior of Q . Then $Q \subset \bigcup P_k$.*

Proof: Let p be a vertex of P_k that lies in the interior of Q . The matching property, the positive convexity, and the matching property combine to show that the link of v is a circle. Hence, $\bigcup P_k$ contains a neighborhood of p . Let $Q' = \bigcup P_k$. If Q is not a subset of Q' , then we can find some point $q \in \text{int}(Q)$ contained in an edge e of some P_k such that Q' does not contain a neighborhood of q . Since e is not a spanning edge, we can follow along e in one direction until we encounter a vertex of some P_k that lies in the interior of Q . Note that Q' cannot contain a neighborhood of any point on \overline{qp} because e can only intersect other edges at vertices. (This comes from the disjoint interiors hypothesis.) Now we can say that Q' does not contain a neighborhood of p , which is a contradiction. ♠

Now we turn to the case of polyhedra. Let Q be a golden polyhedron and let $P_1, \dots, P_n \subset Q$ be a finite collection of strictly convex golden polyhedra having pairwise disjoint interiors. Suppose also that every polyhedron in sight satisfies Equation 33. This means that slicing the polyhedra by golden horizontal planes results in golden polygons. In this section, we explain the test we use to show that Q is partitioned by $\{P_k\}$.

The Slicing Method Let P_h denote the intersection of P with the horizontal plane $z = h$. As we just mentioned, P_h is a golden polygon for all $h \in \mathbf{Z}[\phi]$. Let S be the union of all the third coordinates of all the vertices of our polyhedra, including Q .

We choose a subset $S^* \subset \mathbf{Z}[\phi]$ that is *interlaced* with S . This means that there is one point of S^* between each pair of consecutive points of S . For each $h \in S^*$, we use the 2 dimensional covering test to show that Q_h is partitioned by $\{(P_k)_h\}$.

To see why this test suffices, consider a vertex v in the polygon partition at height h . We would like to see that the link of v remains a circle as we continuously vary the height of the horizontal slice. Suppose some polygons separate as we increase the height of the slice. Then these same polygons overlap as we decrease the height from h . This cannot happen because of the disjoint interior condition.

As we continuously vary the horizontal plane away from h , every slice we encounter satisfies the criterion for our covering test, namely every interior vertex has a circle link. So, for each open interval $I \in \mathbf{R} - S$, and any $h \in I$, the polygons $\{(P_k)_h\}$ partition Q_k . But this is clearly sufficient to show that $\{P_k\}$ partitions Q .

7.5 Polyhedron Exchange Dynamics

Here we explain how we compute various dynamical quantities related to the polyhedron exchange dynamics.

Computing the Orbit Let $\widehat{\Sigma} = (\mathbf{R}/2\mathbf{Z})$ as usual. Let P_1, \dots, P_{64} be the polyhedra in the partition of $\widehat{\Sigma}$ corresponding to the dynamics. Let V_1, \dots, V_{64} be the corresponding vectors associated to each of these regions, so that $\widehat{\Psi}(p) = p + V_k$ for $p \in P_k$.

Here we describe one step of the polyhedron exchange dynamics. Given a point $p \in \widehat{\Sigma}$, we use the interior point detection algorithm to find k such that $p \in P_k$. We then replace p by $p + V_k$ and (if necessary) subtract even integers from the coordinates of $p + V_k$ until the result p' lies in the chosen fundamental domain for $\widehat{\Sigma}$. The map $p \rightarrow p'$ is one step of the polyhedron exchange dynamics.

In practice, we use floating point arithmetic to produce the *candidate* itinerary for a given point or region. By this, we mean the sequence of regions entered by the orbit of that point or region. Once we have the candidate itinerary, we switch to golden arithmetic to verify that the itinerary really is the correct one, as explained below.

There is one fine point of our computer calculations which seem worth mentioning. The map Θ from the Compactification Theorem involves division by 2. Since we want to work entirely in $\mathbf{Z}[\phi]$ we scale all vectors in \mathbf{R}^3 by a factor of 2 when we perform the golden arithmetic calculations. We mention this for the benefit for a reader who wants to survey our computer code.

Domain Verification: Let P_1, \dots, P_{64} and V_1, \dots, V_{64} be as above. Suppose that W is a vector and $I = \{i_1, \dots, i_k\}$ is a sequence, with each i_j in the set $\{1, \dots, 64\}$. We say that the pair (W, I) is *feasible* if

- W lies in the closure of P_{i_1} .
- $W + V_{i_1}$ lies in the closure of P_{i_2} .
- $W + V_{i_1} + V_{i_2}$ lies in the closure of P_{i_3} .
- And so on.

Suppose that Q is a polyhedron with vertices W_1, \dots, W_k . If (W_i, I) is feasible for all $i = 1, \dots, k$, then the first N iterates of $\widehat{\Psi}$ are defined on all interior points of Q and I is the itinerary for such points. This works for the following reason. Since Q lies in the closure of P_{i_1} , each interior point lies in the interior of P_{i_1} . Since $Q + V_{i_1}$ lies in the closure of P_{i_2} , every interior point of $Q + V_{i_1}$ lies in the interior of P_{i_2} . And so on.

Maximality Verification: Suppose we are given a polyhedron Q and some integer N . Suppose that we have already verified that the first N iterates are completely defined on the interior of Q . We would like to check that the first N iterates are nowhere defined on ∂Q .

We perform the following test for each face F of Q . Let W be some point in the interior of Q , as described above. We consider the orbit

$$W_1 = W; \quad W_k = W_{k-1} + V_{i_k}.$$

and check that there is some $k < N$ such that $W_k \in \partial P_k$.

7.6 Tile Creation

Suppose that $p \in \widehat{\Sigma}$ is some point and N is some integer. Here we explain how we produce the maximal convex polyhedron Q on which the first N iterates of $\widehat{\Psi}$ are defined. From the viewpoint of making a rigorous verification, it doesn't matter how we produce Q . Once we have produced Q *somehow*, we use the domain verification algorithm to prove we are correct. For what it is worth, we explain without proof how matter how we produce Q .

We create Q in 4 steps.

1. We produce the first N iterates of p using floating point calculations, as above.
2. We examine how each point of the orbit sits inside polygon of $\widehat{\Sigma}$ that contains it. In the end, we produce a list of numbers, called the *tile data*, which carries the relevant information.
3. We construct a polyhedron Q' from the initial point p and from the tile data.
4. We replace the floating point coordinates of Q' by the best golden real approximations. The result is Q .

We emphasize that there is no guarantee that our construction will produce the correct answer. It could fail at any stage. However, in practice, it works just fine, as later verified by golden arithmetic calculations.

The second and third steps of the sketch above need more explanation.

Pseudo Distances There is a list of 18 planes Π_1, \dots, Π_{18} such that each face of each P_k is parallel to one of these planes. We have precomputed a certain list N_1, \dots, N_{18} of golden vectors such that Π_k is perpendicular to N_k for all k . Given a polygonal face F of one of the 64 polygons P_m , and a point $q \in \mathbf{R}^3$, we define the *pseudo distance* from q to F to be the quantity

$$\delta(q, F) = N_k \cdot (q - V). \quad (66)$$

Here k is such that F is parallel to Π_k , and V is a vertex of F . The quantity $\delta(q, F)$ is independent of the choice of V . We choose the vectors $\{N_k\}$ so that p lies in the interior of P_m iff $\delta(q, F) > 0$ for all F of P_m . This is another way of expressing our interior point detection test.

Tile Data: Let p_1, \dots, p_N be the forward iterates of p . There is a list of 18 planes Π_1, \dots, Π_{18} such that each face of each P_k is parallel to one of these planes. Initially, let D be the length 18 list $\{100, \dots, 100\}$. (The number 100 here is just some convenient large number.) At the k th step of our process, we compute the pseudo-distance from p_k to each of the faces in V_{i_k} . If any of the numbers we get is smaller than the corresponding number in the list D , we replace this number of D by the new, smaller one. When we have done all N steps of the process, we have a final list D . We call D the *tile data*.

Tile Construction: Let D be the tile data associated to the pair (p, N) . Call an index i *active* if $D_i < 100$. If i is active, it means that some face of some P_{i_k} is parallel to Π_i . For each active index, we let Π'_i be the plane parallel to Π_i such that the pseudo-distance from p to Π'_i is D_i . Let H'_i be the half-plane which contains p and has Π'_i as boundary. Then Q' is the intersection of all the H'_i .

In practice, we compute the intersection $\bigcup H_i$ as follows. For each triple of indices (i, j, k) we find the point $\Pi'_i \cap \Pi'_j \cap \Pi'_k$. We call such a point *extraneous* if some Π'_m separates it from the origin. Once we eliminate all extraneous intersections, the remaining ones are the vertices of Q' .

7.7 Tile Filling

Our proof of the Renormalization Theorem involves partitioning the sets \hat{A} and \hat{B} into atoms. From the point of view of giving a proof, it doesn't matter how we produce these partitions. We simply have to verify that they work. However, it seems worthwhile describing what we actually did.

Generating the Individual Atoms: We generate the \hat{B} atoms in the following way. Given a point $p \in \hat{B}$, we compute the points $\hat{\Psi}^k(p)$ for $k = 1, 2, 3, \dots$ until the orbit returns to \hat{B} . Then we use the tile creation algorithm discussed in the previous section.

We do not actually create the \hat{A} atoms by this method. Once we have created all the \hat{B} atoms, we use the action of \hat{R} , the renormalization map, to define the \hat{A} atoms. During the proof of the Renormalization Theorem we then use the domain verification method discussed in §7.5 to verify rigorously that the polyhedra we have defined really are \hat{A} -atoms.

Probing Vectors: We found by trial and error a list of 50 vectors having the following property. Let P be a special polyhedron and let v be a vertex of P . Then at least one of our 50 vectors, when based at v , points into P . Given that our polyhedron exchange map is defined by translations, the same property will hold with respect to any polyhedron – e.g. a periodic tile or an atom – produced by the dynamics. We call these vectors the probing vectors.

Probing a List of Atoms: Suppose we have some list of \hat{B} -atoms contained in some branch Q (i.e. one of the 4 big convex polyhedra in each layer) of \hat{B} . We choose some vertex v of one of the \hat{B} -atoms, or of Q . Let P_1, \dots, P_k be all the atoms on our list that have v as a vertex. Say that a *new probe* is a probing vector W that does not point into and of the P_j .

Given a new probe W , we choose a point v' so that $v' - v$ is a small multiple of W , and then we find the \hat{B} -atom that contains v' and add it to our list.

Generating the Complete List: For each branch, we start with the empty list of \hat{B} atoms. (The initial list of vertices is just the list of vertices of the branch.) We then iterate the probing algorithm, cycling through all vertices repeatedly, until we can find no new probes. This gives us our list of atoms filling the branch. We do this for all branches, and this gives us our complete list of \hat{B} -atoms.

7.8 Variations on a Theme

In §7.5 and §7.6 we explained some constructions that we perform relative to our polyhedron exchange map. In the next chapter, we will do similar things with respect to the square outer billiards map, and also with respect to an auxilliary map which we call the pinwheel map. We will not describe the corresponding operations in detail, because they are so similar, but we will give one example.

Recall that Ψ is the first return map to the strip Σ , as described in §4.2. Suppose we want to verify that a given convex polygon Q is a maximal domain on which the first return map Ψ is entirely defined, and that the return takes n steps. We first use floating point arithmetic to compute a (candidate) length n itinerary I for the center of mass of Q . We then check that (V, I) is feasible, in the sense of §7.5, for all vertices V of Q . Next, we check that $\psi^n(V) \subset \Sigma$ for each vertex V of Σ . Finally, assuming that Q has k edges, we generate a list of k points v_1, \dots, v_k , such that v_j lies in the interior of the j th edge of P and for each j there is some $m = m_j \in \{1, \dots, n-1\}$ such that $\psi^m(v_j)$ lies in the boundary of the relevant domain of definition for ψ .

8 The Pinwheel Lemma

8.1 The Pinwheel Strips

We are interested in the first return map $\Psi : \Sigma_{\pm} \rightarrow \Sigma_{\pm}$ discussed in §4.2. Our Pinwheel Lemma below equates the map Ψ with a new map Π that is simpler to manage. As a step in the Compactification Theorem, we prove a result which we call the Pinwheel Lemma. The Pinwheel Lemma equates Ψ with a map Π which is easier to analyze and “compactify”.

We first explain how to associate 4 pairs (S_i, V_i) for $i = 0, 1, 2, 3$ to the kite K . Here S_i is an infinite strip and V_i is a vector that points from one component of ∂S_i to the other.

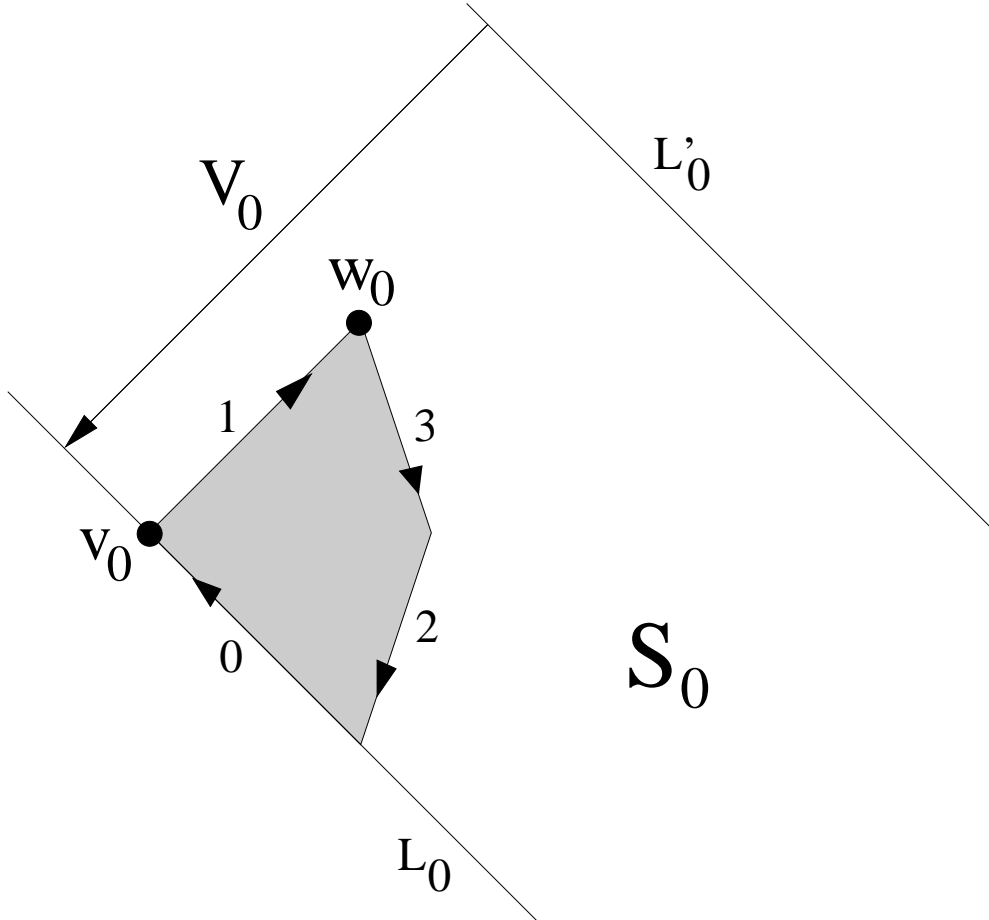


Figure 8.1: The strip associated to e_0 .

We label the edges of K as in Figure 8.1. We call these edges e_0, e_1, e_2, e_3 . We orient these edges so that they go clockwise around K . The labels are such that the lines extending the edges hit the “circle at infinity” (in the projective plane that compactifies \mathbf{R}^2) in counterclockwise cyclic order.

To the edge e_j we associate the strip S_k that has the following properties. One boundary component of L_k of S_k extends e_k . The other boundary component is L'_k is such that the vertex w_k of K farthest from L_k lies halfway between L_k and L'_k . So, K extends exactly up to the centerline of S_k . We define

$$V_k = \pm 2(v_k - w_k). \quad (67)$$

Here v_k is the head vertex of e_k . We choose the signs so that the vectors are as shown in Figure 8.2 below. Figure 8.1 shows S_0 and V_0 .

Let $A = \phi^{-3}$. Here are the formulas for these strip pairs.

- $V_0 = (2, 2)$ and S_0 is bounded by the lines $x + y + 1 \in \{0, 4\}$.
- $V_1 = (0, 4)$ and S_1 is bounded by the lines $x - y - 1 \in \{0, 4\}$.
- $V_2 = (2, -2)$ and S_2 is bounded by the lines $x - Ay - A \in \{0, -4/\phi\}$.
- $V_3 = (-2 - 2A, 0)$ and S_3 is bounded by the lines $x + Ay - A \in \{0, -4/\phi\}$.

Figure 8.2 shows a rough picture of the strips and vectors relative to $K(1/4)$. The picture is very close to what one sees for $K(\phi^{-3})$. The black kite in the middle is $K(1/4)$. Figure 8.2 also shows the strip $\Sigma = \mathbf{R} \times [-2, 2]$, as well as the regions Σ_+ and Σ_- . These regions are denoted $(+)$ and $(-)$ respectively.

To make our later definitions cleaner, we define recursively define

$$V_{4+i} = -V_i; \quad S_{4+i} = S_i. \quad (68)$$

This gives us pairs (S_i, V_i) for all $i = 0, 1, 2, 3, 4, \dots$. The strips repeat with period 4 and the vectors repeat with period 8. However, note that (S, V) and $(S, -V)$ define the same strip maps. So, the strip maps repeat with period 4. Nonetheless, it is useful to distinguish between V_1 and $V_5 = -V_1$, etc. This way, the vectors V_1, \dots, V_8 can be drawn so that they naturally circulate counterclockwise around K , as in Figure 8.2.

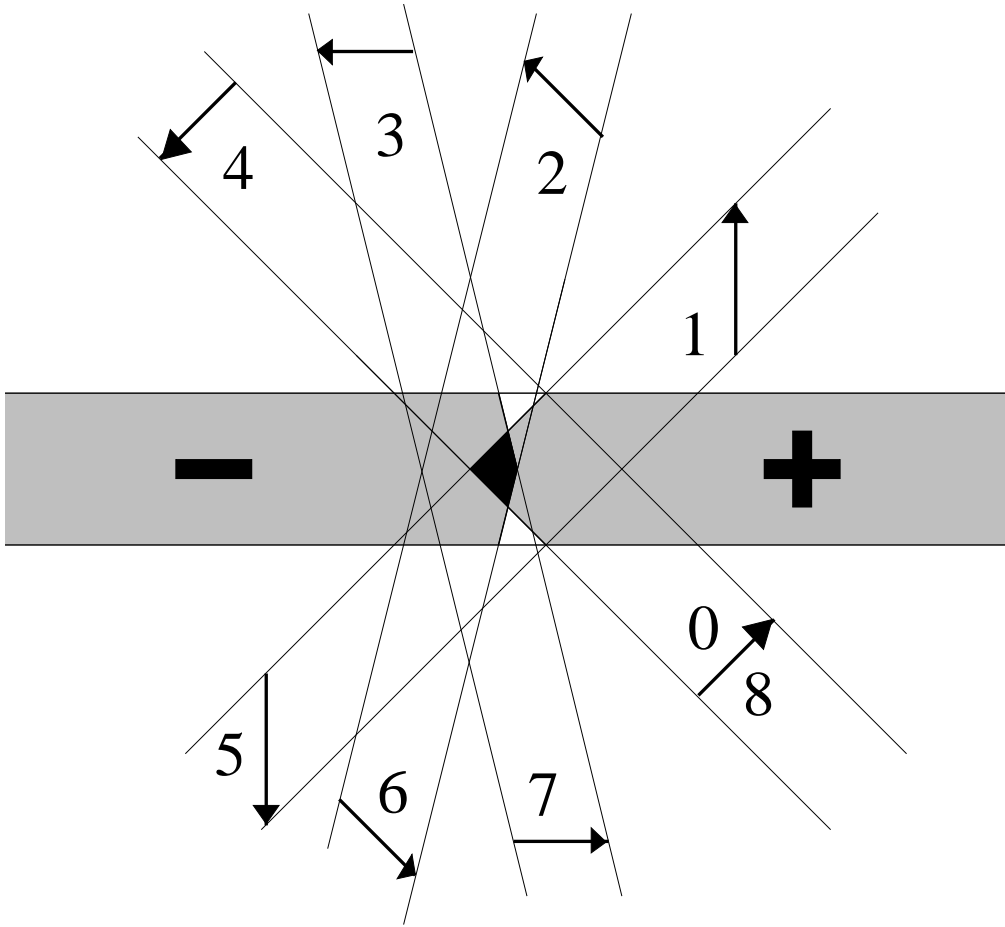


Figure 8.2: The strips and vectors associated to K .

8.2 The Main Result

Let S be an infinite strip in the plane. We say that a *strip pair* is a pair (S, V) where S is a strip and V is a vector that points from one component of ∂S to the other. Given the strip pair, we have an associated *strip map* $E : \mathbf{R}^2 \rightarrow S$, given by the formula

$$E(p) = p + nV \in S. \quad (69)$$

Here $n = n(p)$ is chosen so that $E(p) \in S$. The map E is well-defined on the complement of an infinite discrete set of lines parallel to the boundary components of V . Note that the vectors V and $-V$ define the same map. The pinwheel map is essentially the composition of strip maps.

Referring to the strip pairs (S_i, V_i) defined in the previous section, we recursively define $S_{4+i} = S_i$ and $V_{4+i} = V_i$. This gives us pairs (S_i, V_i) for all integers $i \geq 0$. Of course, these pairs repeat with period 4. We let E_i denote the strip map associated to (S_i, V_i) . This gives strip maps E_1 . We let ζ be the strip map defined relative to the pair $(\Sigma, (0, 4))$. We define the pinwheel map

$$\Pi = \zeta \circ E_8 \circ \dots \circ E_1 : \Sigma \rightarrow \Sigma. \quad (70)$$

Recall that Ψ is the first return map on Σ_+ and likewise the first return map on Σ_- .

We proved versions of the Pinwheel Lemma in [S1], [S2], and [S3], but unfortunately none of these results gives us the exact statement we need.

Theorem 8.1 (Pinwheel) *A point of $\Sigma_+ \cup \Sigma_-$ has a well-defined Π -orbit if and only if it has a well-defined Ψ -orbit, and $\Pi = \Psi$ on all points which have well defined orbits.*

Let Σ_{\pm}^* denote those points in Σ_{\pm} of the form (x, y) with $|x| \leq 20$. We see by direct inspection that the Pinwheel Lemma holds for all points in $\Sigma_{\pm} - \Sigma_*$. Both maps just circulate the point around the kite, as shown in Figure 8.2. See any of the papers cited above for more details about this.

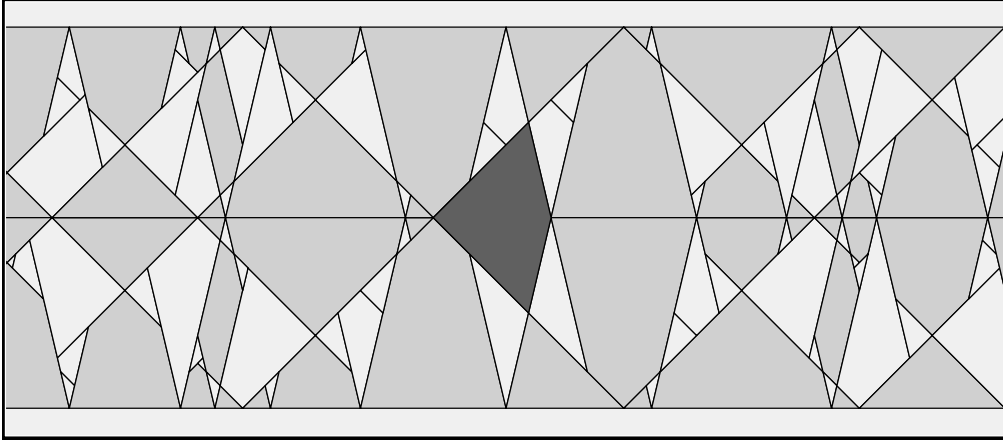


Figure 8.3: Part of the covering

To deal with Σ_{\pm}^* , we make a direct calculation. We produce a list of 572 golden convex polygons P_0, \dots, P_{571} whose union covers Σ_{\pm}^* . Figure 8.3 shows some of these polygons. The central dark polygon is the Penrose Kite.

The two triangles which share the top and bottom vertex with K belong to $\Sigma - \Sigma_{\pm}$, and our dynamical statements do not apply to them. We do not consider them to be on our list.

We use the method described in §7.8 to establish the following result for each polygon P on our list.

- Ψ is entirely defined on the interior of P .
- Ψ^2 is nowhere defined on any non-horizontal edge of P .
- Ψ is the identity on the interior of any horizontal edge of P .

As discussed in §7.4, this result implies that our polyhedra have pairwise disjoint interiors. We then use the method in §7.4 to show that the rectangle $[-20, 20] \times [-2, 2]$ is covered by the union of our polygons and the Penrose kite and the two special triangles. This implies that our union of polygons covers $\Sigma_+^* \cup \Sigma_-^*$.

Next, we computationally establish exactly the same results for Π . Finally, we check that $\Pi = \Psi$ for one interior point of each polygon. Since Π and Ψ are entirely defined on each polygon, we see that $\Pi = \Psi$ on the interiors of each of our polygons, and also on the interiors of the horizontal edges. None of the remaining points have well-defined orbits for either map.

8.3 Discussion

For the purposes of making a rigorous proof, it doesn't really matter how we generate our polygons. However, it seems worth saying a word about how we do it. First of all, we use the method described in §7.6, with Ψ in place of $\widehat{\Psi}$, to generate the individual tiles.

To generate the collection of polygons, we let N be the set of points of the form

$$\left(\frac{i}{100} + \epsilon, \frac{j}{100} + \epsilon\right); \quad i = -2500, \dots, 2500; \quad j = 1, \dots, 199. \quad (71)$$

The small “fudge factor” $\epsilon = 10^{-5}$ is present to guarantee that we choose points on which the map Π is defined. For each $p \in N$ we compute $P(p)$, and then we weed out redundancies and remove the ones that are disjoint from Σ_{\pm}^* .

9 The Compactification Theorem

9.1 Discussion

Recall that $\Sigma = \mathbf{R} \times (-2, 2]$ and $\Pi : \Sigma \rightarrow \Sigma$ is the pinwheel map. In light of the Pinwheel Lemma, it suffices to prove the Compactification Theorem for Π in place of Ψ . We first compare the Compactification Theorem with the corresponding results in [S1] and [S2].

Let $\Sigma_1 \subset \Sigma$ denote the union of two horizontal lines $\mathbf{R} \times \{\pm 1\}$. Let $\widehat{\Sigma}_1$ denote the 2-torus slice of the 3-torus $\widehat{\Sigma}$, obtained by intersecting $\widehat{\Sigma}$ with the plane of height 1. See Figure 4.6. The map Π carries Σ_1 to itself.

Our result [S1, Arithmetic Graph Lemma] is equivalent to the restricted version of the Compactification Theorem, when Σ_1 replaces Σ , and $\widehat{\Sigma}_1$ replaces $\widehat{\Sigma}$. However, the result in [S1] has a somewhat different emphasis. There we were concerned with how $\widehat{\Sigma}_1$ controls the structure of the corresponding arithmetic graphs. In [S2, Master Picture Theorem] we proved a result that is equivalent to [S1, Arithmetic Graph Lemma] for all kite parameters. The result in [S2] is stated in the same general terms as the Compactification Theorem.

The proof in [S2] is different than the proof in [S1]. It is more conceptual and also much easier to generalize. In spite of this, we will prove the Compactification Theorem following the ideas in [S1]. We do this partly because we think of this paper as a sequel to [S1], and also because we would like to take the opportunity to revisit that proof and give a cleaner exposition. The proof we give is specially adapted to the Penrose kite, and would not easily generalize.

9.2 The Covering Property

The first order of business is to show that the 64 golden polyhedra defining the map $\widehat{\Psi}$ really do partition the torus $\widehat{\Sigma}$ into strictly convex polyhedra. (This would be the first step in any approach we took to proving the main result.) We use the tests described in §7.3 to show that the polyhedra involved are strictly convex and have pairwise disjoint interiors. We also verify that they satisfy Equation 33. Finally, we use the slicing method described in §7.4 to show that our polyhedra really do partition $\widehat{\Sigma}$.

9.3 The Proof Modulo Computations

Let \widehat{F} be the fundamental domain for $\widehat{\Sigma}$ that we defined in §5.2. Recall that

$$\Pi = \zeta \circ E_8 \circ \dots \circ E_1 \quad (72)$$

where E_1, \dots, E_8 and ζ are the strip maps defined in §8.2. In the next section, we define the following objects.

- An affine map $\widetilde{\Theta} : \mathbf{R}^2 \rightarrow \mathbf{R}^4$. See Eq. 82
- A locally affine map $\widetilde{\zeta} : \mathbf{R}^4 \rightarrow \widehat{\Sigma}$. See Eq. 83.
- An affine embedding $I_t : \widehat{F} \rightarrow \mathbf{R}^4$, for any $t \in \mathbf{Z}[\phi]$. See Eq. 84.
- A set $\widetilde{L}_k \subset \mathbf{R}^4$ of parallel hyperplanes, for $k = 1, \dots, 8$. See Eqs. 85–86
- Piecewise affine maps $\widetilde{E}_1, \dots, \widetilde{E}_8 : \mathbf{R}^4 \rightarrow \mathbf{R}^4$. See Eqs. 87–88.

Associated to E_j is an infinite family L_j of parallel lines, such that E_j is defined exactly in the complement of the lines. At the same time, \widetilde{E}_k is defined precisely in the complement of \widetilde{L}_k . We write $V \sim W$ if $V - W \in (4\mathbf{Z})^4$.

Lemma 9.1 *The following is true.*

1. $\widetilde{\zeta} \circ \widetilde{\Theta} = \Theta \circ \zeta$.
2. $\widetilde{\zeta} \circ T = \widetilde{\zeta}$.
3. $\widetilde{\zeta} \circ I_t = \text{Identity}$ for all t .
4. $\widetilde{E}_1 \circ I_t$ is independent of t .
5. $T(\widetilde{L}_k) = \widetilde{L}_k$.
6. $\widetilde{\Theta}^{-1}(\widetilde{L}_k) = L_k$.
7. $\widetilde{E}_k \circ T \sim \widetilde{E}_k$.
8. $\widetilde{\Theta} \circ E_k \sim \widetilde{E}_k \circ \widetilde{\Theta}$.

Here T is an element of $(4\mathbf{Z})^4$, considered as a map of \mathbf{R}^4 , and $k = 1, \dots, 8$.

For the next result, we interpret $\Theta(p)$ as a point of \widehat{F} . When the point lies in the interior of \widehat{F} , it is the unique representative of $\Theta(p)$. Otherwise, we make some choice amongst the possibilities.

Lemma 9.2 *There is a 4-element set $\Omega \subset \mathbf{Z}[\phi]$ with the following property. For any $p \in \Sigma$ there is some $t \in \Omega$ such that $I_t \circ \Theta(p) \sim \widetilde{\Theta}(p)$.*

Define the following piecewise (locally) affine map.

$$\widetilde{\Pi} = \widetilde{\zeta} \circ \widetilde{E}_8 \circ \dots \circ \widetilde{E}_1 : \mathbf{R}^4 \rightarrow \widehat{\Sigma}. \quad (73)$$

Lemma 9.3 *Let P be a special polygon. $\widetilde{\Pi} \circ I_t$ is everywhere defined and locally affine on the interior of P , but nowhere defined on the boundary of P .*

We prove the above 3 results in the sections following this one.

Lemma 9.4 *Suppose that p lies in the interior of Σ and $\Theta(p)$ lies in the interior of a special polyhedron. Then Π is well-defined on p .*

Proof: Let $t \in \mathbf{Z}[\phi]$ be the value guaranteed by Lemma 9.2, and let $I = I_t$. We define

$$\widetilde{q}_0 = I \circ \Theta(p); \quad \widetilde{p}_0 = \widetilde{\Theta}(p). \quad (74)$$

By Lemma 9.2 we have $\widetilde{p}_0 \sim \widetilde{q}_0$. We inductively define $\widetilde{q}_k = \widetilde{E}_k(\widetilde{q}_{k-1})$ for $k = 1, \dots, 8$. Applying Lemma 9.3 to the open special polygon that contains $\Theta(p)$, we see that the points $\widetilde{q}_1, \dots, \widetilde{q}_8$ are well defined, meaning that $\widetilde{q}_{k-1} \notin \widetilde{L}_k$.

Since \widetilde{L}_1 is $(4\mathbf{Z}^4)$ -invariant and $\widetilde{p}_0 \sim \widetilde{q}_0$, we have $\widetilde{p}_0 \notin \widetilde{L}_1$. Hence, we may define $\widetilde{p}_1 = \widetilde{E}_1(\widetilde{p}_0)$. By Statement 7 of Lemma 9.1, we have $\widetilde{p}_1 \sim \widetilde{q}_1$. Shifting the indices and repeating the same argument 7 more times, we find that we can inductively define points $\widetilde{p}_k = \widetilde{E}_k(\widetilde{p}_{k-1})$ for $k = 1, \dots, 8$. That is, $\widetilde{\Pi}$ is defined on p . Put another way, $\widetilde{p}_{k-1} \notin \widetilde{L}_k$ for $k = 1, \dots, 8$.

Let $p_0 = p$. We have $\widetilde{p}_0 \notin \widetilde{L}_1$. But then $\widetilde{\Theta}(p_0) \notin \widetilde{L}_1$ because \widetilde{L}_1 is $(4\mathbf{Z})^4$ -invariant and $\widetilde{\Theta}(p_0) \sim \widetilde{p}_0$. By Lemma 9.1, we have $p_0 \notin L_1$. Hence E_1 is defined on p_0 . This allows us to define Let $p_1 = E_1(p_0)$. Shifting the indices and repeating the same argument 7 times, we can inductively define the points $p_k = E_k(p_{k-1})$ for $k = 1, \dots, 8$. In short, the composition $E_8 \circ \dots \circ E_1$ is defined on p .

Finally, the set of lines $\mathbf{R} \times (2\mathbf{Z})$ is invariant under all our strip maps. Since p does not lie on these lines, neither does p_8 . Hence ζ is defined on p_8 . Hence, Π is defined on p and $\Pi(p) = \zeta(p_8)$. ♠

Lemma 9.5 $\tilde{\Pi} \circ \tilde{\Theta} = \Pi \circ \Theta$ whenever both sides are defined.

Proof: We write $V_1 \sim V_2$ when $V_1 - V_2 \in (4\mathbf{Z})^4$. Let $p \in \Sigma$. Repeatedly using Statements 7 and 8 of Lemma 9.1, we have

$$\tilde{E}_8 \circ \dots \circ \tilde{E}_1 \circ \tilde{\Theta}(p) \sim \tilde{\Theta} \circ E_8 \circ \dots \circ E_1(p). \quad (75)$$

By Statement 2 of Lemma 9.1, we get equality above when we apply $\tilde{\zeta}$ to both sides of Equation 75. That is,

$$\tilde{\Pi} \circ \tilde{\Theta}(p) = \tilde{\zeta} \circ \tilde{\Theta} \circ E_8 \circ \dots \circ E_1(p) =^* \Theta \circ \zeta \circ E_8 \circ \dots \circ E_1(p) = \Theta \circ \Pi(p). \quad (76)$$

The starred equality is Statement 1 of Lemma 9.1. ♠

Lemma 9.6 Suppose p_1, p_2 lie in the interior of Σ and $\Theta(p_1), \Theta(p_2)$ lie in the interior of the same special polyhedron. Then $\Theta \circ \Pi(p_j) - \Theta(p_j)$ is the same for $j = 1$ and $j = 2$.

Proof: Let P be the special polygon whose interior contains $\Theta(p_1)$ and $\Theta(p_2)$.

$$\tilde{\Pi} \circ I_{t_1} \circ \Theta(p_1) - \Theta(p_1) = \tilde{\Pi} \circ I_{t_1} \circ \Theta(p_2) - \Theta(p_2) = \tilde{\Pi} \circ I_{t_2} \circ \Theta(p_2) - \Theta(p_2). \quad (77)$$

Here $t_j = t(p_j)$, in the sense of Lemma 9.2, and the subtraction takes place in the abelian group $(\mathbf{R}/2\mathbf{Z})^3$. The first equality comes from the fact that $\tilde{\Pi} \circ I_{t_1}$ is locally affine and entirely defined on P . The second equality has the following explanation. By Statement 4 of Lemma 9.1, the map $\tilde{E}_1 \circ I_t$ is independent of t . Hence $\tilde{\Pi} \circ I_t$ is independent of t .

Noting that $\zeta(p_j) = p_j$, we use the identities in Lemmas 9.1 and 9.5 to conclude that

$$\begin{aligned} \Theta \circ \Pi(p_j) - \Theta(p_j) &= \Theta \circ \Pi(p_j) - \Theta \circ \zeta(p_j) = \tilde{\Pi} \circ \tilde{\Theta}(p_j) - \tilde{\zeta} \circ \tilde{\Theta}(p_j) = \\ &= \tilde{\Pi} \circ I_{t_j} \circ \Theta(p_j) - \tilde{\zeta} \circ I_{t_j} \circ \Theta(p_j) = \tilde{\Pi} \circ I_{t_j} \circ \Theta(p_j) - \Theta(p_j), \end{aligned} \quad (78)$$

Combining Equations 77 and 78, we get the equation we seek. ♠

Recall that $\widehat{\Psi}$ is a piecewise translation on $\widehat{\Sigma}$. For each of the 64 polyhedra P that partition $\widehat{\Sigma}$, we produce a point $p \in \Sigma$ such that

$$\Theta(p) \in \text{interior}(P); \quad \Theta \circ \Pi(p) = \widehat{\Psi}(p). \quad (79)$$

Indeed, this is how we defined $\widehat{\Psi}$ in the first place. If $q \in \Sigma$ is any other point such that $\Theta(q)$ lies in the interior of P then, by Lemma 9.6 we have

$$\Theta \circ \Pi(q) - \Theta(q) = \Theta \circ \Pi(p) - \Theta(p) = \widehat{\Psi}(p) - p = \widehat{\Psi}(q) - q.$$

This shows that $\Theta \circ \Pi = \widehat{\Psi} \circ \Theta$ for all points $p \in \Sigma$ such that $\Theta(p)$ lies in the interior of a special polyhedron.

To finish the deduction of the Compactification Theorem, we just have to understand what happens in exceptional cases.

Lemma 9.7 *Suppose that p lies in the interior of Σ and $\Theta(p)$ does not lie in the interior of a special polygon. Then Π is not defined on p .*

Proof: We will revisit the proof of Lemma 9.4 and use the notation set up in that proof. We will suppose that Π is defined on p and then derive a contradiction. Since Π is defined on p , the compositions E_1 , $E_2 \circ E_1$, etc. are defined on p . Hence, the points p_1, \dots, p_8 are well defined. But then the same argument as in the proof of Lemma 9.4 (but done in reverse) shows that the points $\widetilde{p}_1, \dots, \widetilde{p}_8$ are all defined. Hence, the maps $\widetilde{E}_1, \dots, \widetilde{E}_8$ are all defined on \widetilde{E}_0 . Finally, the map $\widetilde{\zeta}$ is everywhere defined on \mathbf{R}^4 . Hence $\widetilde{\Pi}$ is defined on \widetilde{p}_0 , a point in the boundary of \widetilde{P}_0 . This contradicts the second statement of Lemma 9.3. ♠

It remains only to deal with the points on $\partial\Sigma$. The only issue is the action of the map ζ , and this only comes up at the end of our proof of Lemma 9.4. The problem is that ζ is undefined on the set $\mathbf{R} \times 2\mathbf{Z}$. These lines are invariant under the action of E_1, \dots, E_8 . Hence, if $p \in \partial\Sigma$, then $p_8 \in \mathbf{R} \times 2\mathbf{Z}$ and ζ is not defined on p_8 . However, recall that $\Sigma = (-2, 2] \times \mathbf{R}$. The bottom boundary is left off. Given this convention, we define

$$\zeta(p_8) = (x_8, 2) \in \partial\Sigma. \quad (80)$$

Here x_8 is the first coordinate of p_8 . Once we make this definition, our proofs above go through without a hitch.

9.4 Proof of Lemma 9.1

Recall that

$$\Theta(x_1, x_2) = \left[\left(1, \frac{1}{2}, 0\right) + \left(\frac{x_1}{\phi}, \frac{x_1 - x_2}{2}, x_2\right) \right]_2 \quad \zeta(x_1, x_2) = (x_1, [x_2]_4). \quad (81)$$

Here $[V]_2$ means that we take the image of V in $(\mathbf{R}/2\mathbf{Z})^3$, and $[x_2]_4$ means representative of x_2 in $\mathbf{R}/4\mathbf{Z}$ that lies in $(-2, 2]$.

We define

$$\tilde{\Theta}(x_1, x_2) = \left(x_1 + x_2, x_1 - x_2, \frac{x_1 + x_2}{\phi}, \frac{x_1 - x_2}{\phi} \right). \quad (82)$$

This is exactly the same map we used in [S1]. Next, we define

$$\tilde{\zeta}(x_1, x_2, x_3, x_4) = \left(1, \frac{1}{2}, 0\right) + \frac{1}{2}(x_3 + x_4, x_2, x_1 - x_2) \in (\mathbf{R}/2\mathbf{Z})^3. \quad (83)$$

For any $t \in \mathbf{Z}[\phi]$ we define We define

$$I_t(x_1, x_2, x_3) = (2y_2 + y_3, 2y_2 + y_3, y_1, y_1) + (y_3, -y_3, \frac{y_3}{\phi} - t, -\frac{y_3}{\phi} + t),$$

$$(y_1, y_2, y_3) = (x_1, x_2, x_3) - \left(1, \frac{1}{2}, 0\right). \quad (84)$$

Now we define the families \tilde{L}_k of parallel hyperplanes. We introduce the functions $g_k : \mathbf{R}^4 \rightarrow \mathbf{R}$ as follows.

$$g_1 = x_2 + 1; \quad g_2 = x_2 + x_3 + 2 + \phi; \quad g_3 = x_1 + x_4 + 2 + \phi; \quad g_4 = x_1 + 1. \quad (85)$$

We define

$$\tilde{L}_{k+4} = \tilde{L}_k = g_k^{-1}(4\mathbf{Z}). \quad (86)$$

Statements 1-4:

1. A direct calculation shows $\tilde{\zeta} \circ \tilde{\Theta} = \Theta$ and $\Theta = \Theta \circ \zeta$. Hence $\tilde{\zeta} \circ \tilde{\Theta} = \Theta \circ \zeta$.
2. It is immediate from the formula that $\tilde{\zeta} \circ T = \tilde{\zeta}$ for any $T \in (4\mathbf{Z})^4$.
3. A direct calculation shows that $\tilde{\zeta} \circ \tilde{I}_t$ is the identity.
4. It is immediate from the equations that \tilde{L}_k is $(4\mathbf{Z}^4)$ -invariant.

For $x \in \mathbf{R} - 4\mathbf{Z}$ let $\tau(x)$ denote the greatest element in $4\mathbf{Z}$ that is less than x . Define

$$\gamma_k(X) = \phi \times (\tau \circ g_k(X)) \in 4\phi\mathbf{Z}. \quad (87)$$

The (\times) in Equation 87 means multiplication. Here $X = (x_1, x_2, x_3, x_4)$ is a point of $\mathbf{R}^4 - \tilde{L}_k$. Finally, we define

$$\begin{aligned} \tilde{E}_1 &= \begin{pmatrix} x_1 \\ x_2 \\ -x_2/\phi + x_3 + x_4 \\ x_2/\phi \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \gamma_1 \\ -\gamma_1 \end{pmatrix} \\ \tilde{E}_2 &= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ -x_1 + x_2/\phi + x_3\phi \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\gamma_2 \end{pmatrix} \\ \tilde{E}_3 &= \begin{pmatrix} x_1\phi - x_2 - x_3 + x_4\phi \\ x_1/\phi + 0x_2 - x_3 + x_4\phi \\ -x_1/\phi + x_2 + 2x_3 - x_4\phi \\ -x_1/\phi + x_2 + x_3 - x_4/\phi \end{pmatrix} + \begin{pmatrix} -\gamma_3 \\ -\gamma_3 \\ \gamma_3 \\ \gamma_3 \end{pmatrix} \\ \tilde{E}_4 &= \begin{pmatrix} x_1 \\ x_2 \\ x_1/\phi \\ x_4 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -\gamma_4 \\ 0 \end{pmatrix} \end{aligned} \quad (88)$$

Here we have simplified our notation slightly by writing $\tilde{E}_1 = \tilde{E}_1(X)$ and $\gamma_1 = \gamma_1(X)$, etc. Each \tilde{E}_k is the sum of a linear transformation and a piecewise constant vector-valued function. The linear part is defined on all of \mathbf{R}^4 , but the vector-valued function changes value when one passes through a hyperplane of \hat{L}_k . We define $\tilde{E}_{4+k} = \tilde{E}_k$.

Statement 5: The maps I_0 and I_t agree up to post-composition with the translation form $V \rightarrow V + (0, 0, -t, t)$. Inspecting the formula for \tilde{E}_1 , we see that the only expression involving the third and fourth coordinates is $x_3 + x_4$. Hence $\tilde{E}_1 \circ \tilde{I}_t$ is independent of t . Hence $\tilde{\Pi} \circ I_t$ is independent of t .

Statement 6: We can deduce the formulas for the lines $L_k \subset \mathbf{R}^2$ by looking at the formulas for the strip pairs given in §8.1. We have $L_{4+k} = L_k$, so we just need to consider L_k for $k = 1, 2, 3, 4$. Let $A = \phi^{-3}$.

- L_1 consists of those lines satisfying $x_1 - x_2 - 1 \in 4\mathbf{Z}$.
- L_2 consists of those lines satisfying $x_1 - Ax_2 - A \in 4\phi^{-1}\mathbf{Z}$.
- L_3 consists of those lines satisfying $x_1 + Ax_2 - A \in 4\phi^{-1}\mathbf{Z}$.
- L_4 consists of those lines satisfying $x_1 + x_2 + 1 \in 4\mathbf{Z}$.

The fact that $\tilde{\Theta}^{-1}(\tilde{L}_k) = L_k$ is obvious for $k = 1, 4$. For $k = 2$ we compute

$$g_2 \circ \tilde{\Theta}(x_1, x_2) = \phi x_1 - \phi^{-2}x_2 + 4 - \phi^{-2} = \phi(x_1 - Ax_2 - A).$$

The result is obvious from here. The proof for $k = 3$ is similar.

Statement 7: Let e_1, e_2, e_3, e_4 be the standard basis vectors of \mathbf{R}^4 . let

$$u(j, k, X) = \tilde{E}_j(X + 4e_k) - \tilde{E}_j(X). \quad (89)$$

Statement 7 is equivalent to the statement that $u(j, k, X) \in (4\mathbf{Z})^4$ for all j, k and X .

We see directly that $u(1, k, X) \in (4\mathbf{Z})^4$ for $k = 1, 3, 4$, and we compute

$$u(1, 2, X) = (0, 0, -4/\phi, 4/\phi) + (0, 0, 4\phi, -4\phi) = (0, 0, -4, 4).$$

This proves the result for \tilde{E}_1 . The proof for \tilde{E}_4 is similar.

We see directly that $u(2, k, X) \in (4\mathbf{Z})^4$ for $j = 1, 4$, and we compute

$$u(2, 2, X) = (0, 0, 0, 4/\phi) + (0, 0, 0, -4\phi) = (0, 0, 0, -4),$$

$$u(2, 3, X) = (0, 0, 0, 4\phi) + (0, 0, 0, -4\phi) = (0, 0, 0, 0).$$

This proves the result for \tilde{E}_2 .

For \tilde{E}_3 , we compute

$$u(3, 1, X) = (4\phi, 4/\phi, -4/\phi, -4\phi) + (-4\phi, -4\phi, 4\phi, 4\phi) = (0, -4, 4, 4)$$

$$u(3, 2, X) = (-4, 0, 4, 4); \quad u(3, 3, X) = (-4, -4, 8, 4).$$

$$u(3, 4, X) = (4\phi, 4\phi, -4\phi, -4/\phi) + (-4\phi, -4\phi, 4\phi, 4\phi) = (0, 0, 4, 4).$$

Statement 8: We want to prove that

$$\tilde{\Theta} \circ E_k \sim \tilde{E}_k \circ \tilde{\Theta},$$

where \sim denotes equivalence mod $(4\mathbf{Z})^4$. Recall that the strip map E_k is defined in terms of a pair (S_k, V_k) , where S_k is a strip and V_k is a vector. The two lines of ∂S_k are consecutive lines of L_k . Hence, by Lemma 9.1, the set $\tilde{S}_k = \tilde{\Theta}(S_k)$ lies between two hyperplanes in \tilde{L}_k . Hence \tilde{E}_k is an affine map on \tilde{S}_k . Indeed, \tilde{E}_k is linear on \tilde{S}_k because

$$\gamma(0, 0, 0, 0) = 0; \quad (0, 0, 0, 0) = \tilde{\Theta}(0, 0) \in \tilde{\Theta}(S_k) = \tilde{S}_k$$

A direct calculation shows that the linear part of \tilde{E}_k is the identity on the 2-plane $\tilde{\Theta}(\mathbf{R}^2)$. Hence \tilde{E}_k is the identity on $\tilde{\Sigma}$. In summary

$$\tilde{\Theta}(p) = \tilde{E}_k \circ \tilde{\Theta}(p) \quad \forall p \in \text{interior}(\Sigma). \quad (90)$$

Let $\tilde{V}_k = \tilde{\Theta}(V_k)$. Given the piecewise affine nature of our maps, the value of $\tilde{E}_k(X + \tilde{V}_k) - \tilde{E}_k(X)$ is independent of the choice of $X \in \mathbf{R}^4 - \tilde{L}_k$. Compare the proof of Statement 7. A direct calculation shows that $\tilde{E}_k(\tilde{V}_k) \in (4\mathbf{Z})^4$. Therefore

$$\tilde{E}_k(X + \tilde{V}_k) \sim \tilde{E}_k(X); \quad \forall X \in \mathbf{R}^4 - \tilde{L}_k.$$

Iterating this formula, we get

$$\tilde{E}_k(X + n\tilde{V}_k) \sim \tilde{E}_k(X); \quad \forall X \in \mathbf{R}^4 - \tilde{L}_k. \quad (91)$$

Here n is an integer.

Every point $p' \in \mathbf{R}^2 - L_k$ has the form $p + nV_k$ for some p in the interior of S_k . Note that $E_k(p') = p$. We have

$$\tilde{\Theta} \circ E_k(p') = \tilde{\Theta}(p) = \tilde{E}_k \circ \tilde{\Theta}(p) \sim$$

$$\tilde{E}_k(\tilde{\Theta}(p) + n\tilde{V}_k) = \tilde{E}_k \circ \tilde{\Theta}(p + nV_k) = \tilde{E}_k \circ \tilde{\Theta}(p').$$

This proves Statement 8.

9.5 Proof of Lemma 9.2

Let $p = (x_1, x_2) \in \Sigma$. When we interpret $\Theta(p)$ as a point of the fundamental domain \widehat{F} , we are choosing integers A_1, A_2, A_3 such that

$$\Theta(p) = \left(2A_1 + 1, 2A_2 + \frac{1}{2}, 2A_3\right) + \left(\frac{x_1}{\phi}, \frac{x_1 - x_2}{2}, x_2\right) \in \widehat{F}. \quad (92)$$

Note that

$$A_3 \in \{0, 1\} \quad (93)$$

because $x_2 \in (-2, 2]$. Using Equation 92 and the definition of the map I_t , we compute

$$I_t \circ \Theta(p) = \widetilde{\Theta}(p) + \left(4A_2 + 4A_3, 4A_2, 2A_1 + \frac{2A_3}{\phi} + t, 2A_1 - \frac{2A_3}{\phi} - t\right). \quad (94)$$

The appropriate choice of t from amongst the values

$$0; \quad 2; \quad -2/\phi; \quad 2 - 2/\phi \quad (95)$$

leads to

$$I_t \circ \Theta(p) \sim \widetilde{\Theta}(p)$$

This proves Lemma 9.2.

9.6 Proof of Lemma 9.3

The arguments we give here are very similar to what we discussed in §7.5, but the setting is different enough that we will go through the details.

Given the 64 special polyhedra P_1, \dots, P_{64} , we define the 256 polyhedra

$$P_{ij} = I_{t_i}(P_j) \subset \mathbf{R}^4; \quad i \in \{1, 2, 3, 4\} \quad j \in \{1, \dots, 64\}. \quad (96)$$

Lemma 9.8 $\widetilde{\Pi}$ is defined on the interior of P_{ij} for all $i = 1, 2, 3, 4$ and all $j = 1, \dots, 64$.

Proof: Let q be the center of mass of P_{ij} . Using floating point arithmetic, we associate to q a certain 8-tuple of integers, as follows. Let $q_0 = q$ and inductively define $q_k = \widetilde{E}_k(q_{k-1})$. We define

$$n_k = \phi^{-1} \gamma_k(p_{k-1}). \quad (97)$$

This integer locates the component of $\mathbf{R}^4 - \tilde{L}_k$ that contains q_{k-1} . The 8-tuple of interest to us is (n_1, \dots, n_8) . For the purposes of our proof, it doesn't matter how we generated the sequence (n_1, \dots, n_8) . This sequence just serves as a guide for our rigorous calculations.

Guided by the sequence we have produced, we define the map \tilde{E}_k^* to be the same map as \tilde{E}_k , except that $4\phi n_k$ replaces γ_k in the formula. We check, using golden arithmetic, that the composition $\tilde{E}_8^* \circ \dots \circ \tilde{E}_1^*$ is defined on all the vertices of P_{ij} . For each such vertex v_0 we inductively define $v_k = \tilde{E}_k^*(v_{k-1})$. We check that

$$g_k(v_k) \in [n_k, n_k + 4]. \quad (98)$$

Here g_k is as in Equation 85.

The same argument we gave when we discussed domain verification in §7.5 now says that $\tilde{E}_8 \circ \tilde{E}_1$ is defined on the interior of P_{ij} . Indeed, we can now say rigorously that the sequence (n_1, \dots, n_8) “works” for all points in the interior of P_{ij} . ♠

Lemma 9.9 *$\tilde{\Pi}$ is not defined on any boundary point of P_{ij} for any indices $i = 1, 2, 3, 4$ and any $j = 1, \dots, 64$.*

Proof: We use the notation from the previous proof. For each face F of P_{ij} we exhibit an index k such that one of the following two equations simultaneously holds relative to any vertex v_0 of F :

- $g_k(v_{k-1}) = n_k$.
- $g_k(v_{k-1}) = n_{k+4}$.

This shows that

$$F_{k-1} = \tilde{E}_{k-1} \circ \dots \circ \tilde{E}_1(F)$$

lies entirely in one hyperplane of \tilde{L}_k . Hence \tilde{E}_k is nowhere defined on F_{k-1} . Hence $\tilde{\Pi}$ is nowhere defined on F . Since F is an arbitrary face, $\tilde{\Pi}$ is nowhere defined on the boundary of P_{ij} . ♠

10 Proof of the Renormalization Theorem

10.1 Some Terminology

We use the notation from the Renormalization Theorem. Here we establish some terminology which will be useful to us in this chapter and in later ones.

Forward Atoms: Recall that \hat{A} is partitioned into atoms. We call these atoms *forward \hat{A} -atoms*. We define the forward \hat{B} -atoms in the same way.

Neutral and Backward Atoms Let P be a forward \hat{A} atom. There is some $n > 1$, depending on P , such that $\hat{\Psi}^k(\hat{P})$ is disjoint from \hat{A} for $k = 1, \dots, n-1$ but $\hat{\Psi}^n(P) \subset \hat{A}$. We call the polyhedra $\hat{\Psi}^k(P)$ *neutral \hat{A} -atoms* when $k = 1, \dots, n-1$. We call $\hat{\Psi}^n(P)$ a *backward \hat{A} -atom*. \hat{A} is also partitioned into backward \hat{A} -atoms, and the first return map of $\hat{\Psi}^{-1}$ is entirely defined and a translation on each one.

Chains: Let P be a forward \hat{A} -atom. We call the sequence $P, \dots, \hat{\Psi}^{n-1}(P)$ a forward \hat{A} -chain. So, each forward \hat{A} -chain is a finite union of \hat{A} -atoms, the first of which is forward and the rest of which are neutral. We make the same definitions for \hat{B} .

Backward Chains: We define *backward chains* just as we defined the forward chains, except that we use backward atoms in place of forward atoms, and we use $\hat{\Psi}^{-1}$ in place of $\hat{\Psi}$. We think of the backward atom as being the last atom in the backwards chain. For each forward chain there is a backward chain that shares all its atoms except the first one. Likewise, for every backward chain there is a forward chain that shares all of its atoms except for its last one.

Types: We distinguish two types of layers of \hat{A} . The *type-1* layers are the ones on which the restriction of \hat{R} to that layer is a translation. The other layers we call type-2. Note that this definition is consistent with the notion of type defined in association with the Reduction Theorems. We define the type of a \hat{A} -atom and a \hat{A} -chain according to the type of the layer that contains it.

10.2 Statements 1, 2, and 5

We compute the list of \hat{A} and \hat{B} atoms using the method described in §7.7. Figure 10.1 shows a slice of the set of \hat{B} -atoms at a particular height.

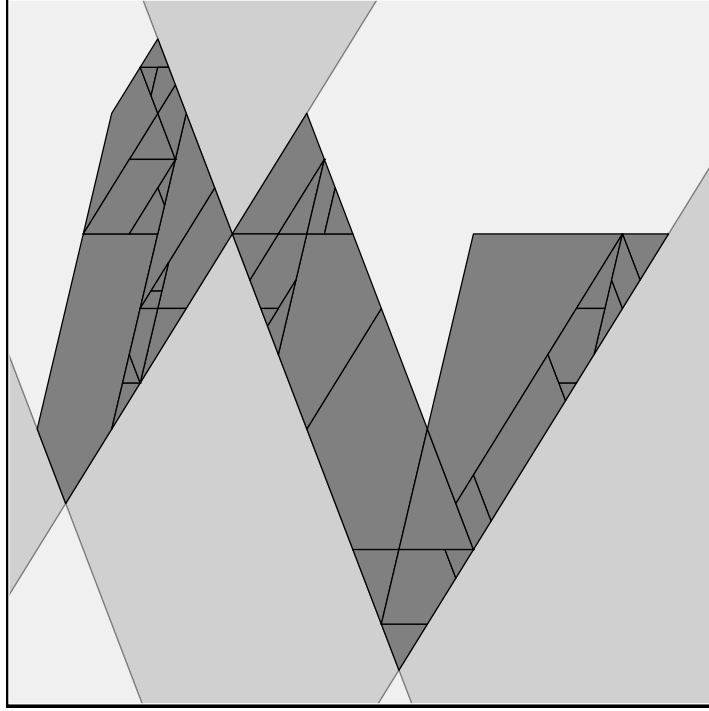


Figure 10.1: The \hat{B} -atoms sliced at $-7 + 4\phi \approx .5278$

We use the method described in §7.5 to check that each of the polyhedra on our candidate list of \hat{A} -atoms is indeed a \hat{A} -atom. We do the same thing for the \hat{B} -atoms. In the case of the \hat{B} -atoms, we also verify that each face F of a \hat{B} -atom has one of 3 properties.

- The first return map $\hat{\Psi}|_{\hat{B}}$ is not defined on F . That is, some small power of $\hat{\Psi}$ maps F into one of the faces of one of the 64 special polyhedra that define $\hat{\Psi}$.
- F lies in $\partial\hat{B}$.
- The first return of F to \hat{B} lies in $\partial\hat{B}$.

Since these properties hold for every edge, all the \hat{B} atoms are maximal.

Remark: At some point we also checked that all the \hat{A} -atoms are maximal, but we did not save the computer code for this. This check is not necessary for our arguments.

Given that the \hat{B} -atoms are all maximal, they have pairwise disjoint interiors. We also check that they are strictly convex. We now apply the slicing method of §7.4 to each branch of each layer of \hat{B} . We conclude that the \hat{B} -atoms partition \hat{B} . That is, we have obtained the complete list.

As we mentioned in §7.7, we obtain the candidate list of \hat{A} atoms by pulling back the \hat{B} atoms by the action of the renormalization map \hat{R} . From this way of constructing things, we automatically know that the \hat{A} -atoms are strictly convex and have pairwise disjoint interiors. (We don't need the maximality for this.) We now apply the slicing method to \hat{A} just as we did for \hat{B} . We conclude that \hat{A} is partitioned into the \hat{A} atoms on our list.

We have set things up so that, on each type-1 layer, \hat{R} establishes a bijection between the forward \hat{A} -atoms and the forward \hat{B} -atoms. We check that this bijection is compatible with action of the first return maps $\hat{\Psi}|_{\hat{A}}$ and $\hat{\Psi}|_{\hat{B}}$. Since we have already verified that everything in sight is an atom, we just need to check the action of the relevant maps on one interior point of each atom. For the check, we choose interior golden points as in §7.3.

We have set things up so that, on each type-2 layer, \hat{R} establishes a bijection between the forward \hat{A} -atoms and the backward \hat{B} -atoms. In the same way as in the type-1 case, we check that the action of \hat{R} is compatible with action of the first return maps $\hat{\Psi}|_{\hat{A}}$ and $\hat{\Psi}^{-1}|_{\hat{B}}$.

These calculations prove Statement 1 of the Renormalization Theorem.

Statement 2 of the Renormalization Theorem is really just a description of the action of \hat{R} . We have constructed \hat{R} so that it acts on the horizontal slices as R acts on $\mathbf{R}/2\mathbf{Z}$.

Statement 5 is purely combinatorial. We simply check, in each case, that each \hat{B} -chain has a shorter length than the corresponding \hat{B} -chain.

Remark: For the record, we mention that there are 678 \hat{B} atoms and $3 \times 678 = 2034$ \hat{A} -atoms. We also compute that the longest \hat{A} -chain has length 703 and the longest \hat{B} -chain has length 109. In other words, for generic points in \hat{A} , the first return map is defined in at most 704 iterates and for generic points in \hat{B} the first return map is defined in at most 110 iterates.

10.3 Statement 3

When we lift $\Theta(\Sigma)$ to \mathbf{R}^3 we see that it consists precisely in the planes spanned by the vectors

$$(2\phi^{-1}, 1, 0); \quad (0, -2, 1)$$

which intersect the line $\{0\} \times \mathbf{R} \times \{0\}$ in points of the form $(0, t, 0)$ where $t \in \mathbf{Z}[\phi]$. This follows from a routine calculation. Another calculation shows that each choice of \hat{R} mentioned above preserves this set.

10.4 Statements 4 for \hat{B}

Recall that \hat{B} is partitioned into 6 layers, as discussed in §5.4. It turns out that some periodic tiles intersect some layers of \hat{B} but not others. To make our proof go cleanly, we introduce the notation of a \hat{B} -periodic tile. A \hat{B} periodic tile is an intersection of the form

$$P \cap \left((\mathbf{R}/2\mathbf{Z}) \times (\mathbf{R}/2/Z) \times J_\lambda \right) \quad (99)$$

where J_λ is the 6 intervals J_1, \dots, J_6 in the partition defining the layers of \hat{B} .

Using essentially the same method as in §8.3, we generate a candidate list of \hat{B} periodic tiles whose orbits avoid \hat{B} . In practice, we simply enumerate all the periodic tiles whose orbits avoid at least one layer of \hat{B} , and then we keep track of which layers each orbit intersects. We check that the period of the longest tile on the list is 33.

Let $\lambda \in \{1, 2, 3, 4, 5, 6\}$ be an integer that describes a layer of \hat{B} currently of interest to us. Let $i \in \{1, \dots, 64\}$ be an index that describes one of the 64 special polyhedra that partition $\hat{\Sigma}$.

Let $\mathcal{B}(i, \lambda)$ be the union of all the \hat{B} -tiles and forward/neutral \hat{B} -atoms that are contained in the region

$$P(i, \lambda) = P_i \cap \left((\mathbf{R}/2\mathbf{Z}) \times (\mathbf{R}/2/Z) \times J_\lambda \right) \quad (100)$$

By construction, these tiles have pairwise disjoint interiors. Say that (i, λ) is *good* if the tiles in $\mathcal{B}(i, \lambda)$ partition the region in Equation 100.

Lemma 10.1 *Statement 4 holds for \hat{B} provided that every pair (i, λ) is good.*

Proof: Let p_1, \dots, p_{109} be a (non-periodic) generic orbit portion of length 109. Consider the point $p = p_{109}$. There is some pair (i, λ) such that $p \in P(i, \lambda)$. Since (i, λ) is good and p is generic, p lies in the interior of either a \widehat{B} periodic tile or in the interior of a forward/neutral \widehat{B} -atom. In the former case, p has period at most 33. This is a contradiction. If p lies in the interior of a forward \widehat{B} -atom, then $p \in \widehat{B}$, by definition, and we are done.

So, suppose p lies in the interior of neutral \widehat{B} -atom. Since the longest \widehat{B} -chain has length 109, there is some $m \leq 109$ such that $\widehat{\Psi}^{-m}(p) \in \widehat{B}$. But then $p_{109-m} \in \widehat{B}$. This covers all the cases. ♠

Our computer code is equipped to verify that all indices are good, but we can get away with doing less computational work. The indices (i, λ) are automatically good for $i \in \{1, \dots, 22\}$. These correspond to the special polyhedra on which $\widehat{\Psi}$ acts trivially. So, we only have to consider indices $i > 22$.

We can eliminate more indices by considering the dynamics of $\widehat{\Psi}$. For instance, $\widehat{\Psi}(P_{26}) = P_{25}$. Hence, if $(25, \lambda)$ is good, then so is $(26, \lambda)$. In Lemma 14.1, we prove that the list of indices

$$I = \{23, 25, 32, 40, 41, 44, 46, 53, 61, 62\} \quad (101)$$

is such that every well-defined and nontrivial orbit intersects P_i for some $i \in I$. For this reason, we only have to check that (i, λ) is good for $i \in I$ and $\lambda \in \{1, 2, 3, 4, 5, 6\}$.

There is one more savings we can make. Everything in sight is invariant under the involution

$$(x, y, z) \rightarrow (2, 2, 2) - (x, y, z). \quad (102)$$

This involution permutes the special polyhedra, commutes with $\widehat{\Psi}$, preserves \widehat{A} and \widehat{B} , and permutes the list of atoms. For this reason, it suffices to prove that the indices (i, λ) are good for $i \in I$ and $\lambda \in \{1, 2, 3\}$. Using the slicing method of §7.4 we check that this is the case. This establishes Statement 4 for \widehat{B} . Incidentally, our calculation show that our candidate list of \widehat{B} -periodic tiles is correct and complete.

Remark: The calculation we make is fairly massive, but the reader can use our program and see the individual slices plotted as they are tested. The reader can also select any index (i, λ) and plot the corresponding partition.

10.5 Statement 4 for \hat{A}

In principle, we could prove Statement 4 for \hat{A} just as we proved it for \hat{B} . This time there are 18 layers. The problem with this approach is that it leads to a massive calculation. For instance, the partition associated to the index pair $(23, 7)$ has 2562 polyhedra.

Were we to take this approach (and succeed) we would know than any generic orbit portion of length 703 intersects \hat{A} . The bound of 703 is the sharp constant. There is an orbit portion of length 702 that avoids \hat{A} . We will take a different approach and get the slightly worse constant 812. The benefit we get from the other approach is that the calculation we make is vastly shorter. The idea is to use the result we have already proved for \hat{B} and only calculate “the difference” between \hat{B} and \hat{A} , so to speak.

Recall that each layer of \hat{B} is divided into 4 branches. Moreover, each layer of \hat{B} is partitioned into finitely many smaller pieces, corresponding to the layers of \hat{A} . We call these pieces of \hat{B} the *sub-layers* of \hat{B} . Precisely,

- Layer 1 of \hat{B} has 5 sublayers.
- Layer 2 of \hat{B} has 1 sublayers.
- Layer 3 of \hat{B} has 3 sublayers.
- Layer 4 of \hat{B} has 3 sublayers.
- Layer 5 of \hat{B} has 1 sublayers.
- Layer 6 of \hat{B} has 5 sublayers.

The palindromic nature of the list comes from the involutive symmetry mentioned in the previous section.

Each layer of \hat{B} is divided into 4 branches. Each branch is a golden polyhedron. By slicing each branch with relevant horizontal planes, we can say that each sublayer of \hat{B} is also divided into 4 branches. In summary, \hat{B} has 18 sublayers and each sublayer has 4 branches. We index the sublayers by $\lambda \in \{1, \dots, 18\}$ and the branches by $\beta \in \{1, 2, 3, 4\}$.

Finally we discuss \hat{A} . We define the \hat{A} -periodic tiles just as we defined the \hat{B} -periodic tiles. We find a candidate list of all \hat{A} -periodic tiles as we did in the \hat{B} case. The longest period in this case is 213. We say that a pair (λ, β) is *good* if the polyhedron corresponding to (λ, β) is partitioned into a union of \hat{A} -periodic tiles and forward/neutral \hat{A} -atoms.

Lemma 10.2 *Statement 4 holds for \hat{A} provided that every index (λ, β) is good.*

Proof: the same argument as in Lemma 10.1 proves that any orbit portion of length 703 that starts in \hat{B} intersects \hat{A} . Combining this result with Lemma 10.1, we see that every generic orbit portion of length $812 = 703 + 109$ intersects \hat{A} . ♠

To finish the proof of Statement 4 for \hat{A} we just have to prove that every index (λ, β) is good. Using the involutive symmetry discussed in the previous section, it suffices to prove this result for indices $\lambda \leq 9$. Using the slicing method of §7.4 we verify that the indices (λ, β) are good for $\lambda \in \{1, \dots, 9\}$ and $\beta \in \{1, 2, 3\}$. This completes the proof.

Remark: This time, our calculations do not verify that our list of periodic \hat{A} tiles is complete. Rather, all we know is that each polyhedron on the list is indeed a \hat{A} -periodic tile, and that our list of \hat{A} -periodic tiles whose orbits intersect \hat{B} is complete. We can still say something useful, even with an incomplete list. Statement 4 for \hat{A} implies that $\hat{\Sigma}$ is partitioned into the forward/neutral \hat{A} -atoms and a finite union of \hat{A} periodic tiles, some of which perhaps are not on our list. If we have a generic orbit portion that is not contained in a periodic orbit, then it must intersect one of the forward or neutral \hat{A} -atoms. But the longest \hat{A} -chain has length 703. this establishes that a generic orbit portion of length 703 intersects \hat{A} provided it does not lie on a periodic orbit. The constant 703 is sharp because of the \hat{A} -chain of length 703.

10.6 Statement 6

Any generic orbit that avoids \hat{B} is contained in the orbit of one of the \hat{B} -periodic tiles on our list. We check that all these orbits avoid \hat{A} . Hence, any orbit that avoids \hat{B} also avoids \hat{A} .

11 Critical Renormalization Calculations

11.1 Discussion

The goal of this chapter is to prove the Fundamental Orbit Theorem and a second result which we call the Fixed Point Theorem. Both results involve fixed points of the renormalization operator. They are the cleanest applications of the Renormalization Theorem. In the next chapter, we will give some messier applications, including the proof of the Near Reduction Theorem.

We first discuss the general nature of our calculations. We use the notation from the previous chapter. Let \mathcal{A} be the set of \widehat{A} -chains. Likewise define \mathcal{B} . The Renormalization Theorem gives us a canonical 3-to-1 map from \mathcal{A} to \mathcal{B} . This map does not depend on any conventions concerning the boundaries of the layers of \widehat{A} . Let $\chi : \mathcal{A} \rightarrow \mathcal{B}$ denote the map just described.

Since Σ can be considered as a dynamically invariant subset of $\widehat{\Sigma}$, we can transfer the structure to Σ . For instance, an *A-atom* of Σ is a connected intersection of Σ with the \widehat{A} -atoms. Every polygon in sight is a golden polygon. We deal with these polygons and polyhedra using the methods from §6.

Given an *A-atom* X and a *B-atom* Y , we write $X \rightarrow Y$ if $\chi([X]) = [Y]$. Here $[X]$ is the forward *A-chain* containing X and $[Y]$ is the *B-chain* containing Y . We take the *B-chains* to be forwards when they are discussed in connection with type-1 *A-chains*, and backwards when discussed in connection with type-2 *A-chains*.

Important Observation: By construction, a generic point never intersects the boundary of an *A-atom* or the boundary of a *B-atom*.

11.2 Proof of the Fundamental Orbit Theorem

Referring to the notation in the Fundamental Orbit Theorem, we find a partition of each T_{ij}^\pm by a union of *A-atoms* and periodic tiles such that

$$\tau \rightarrow \psi^{-1}(R_{ij}^\pm(\tau)), \quad (103)$$

for every *A-atom* τ . The periodic tiles all lie in $\mathcal{T}(2)$.

For future reference, we refer to the above mentioned partitions as the *A-partitions*. The left side of Figure 11.1 shows the partition for T_{11}^+ .

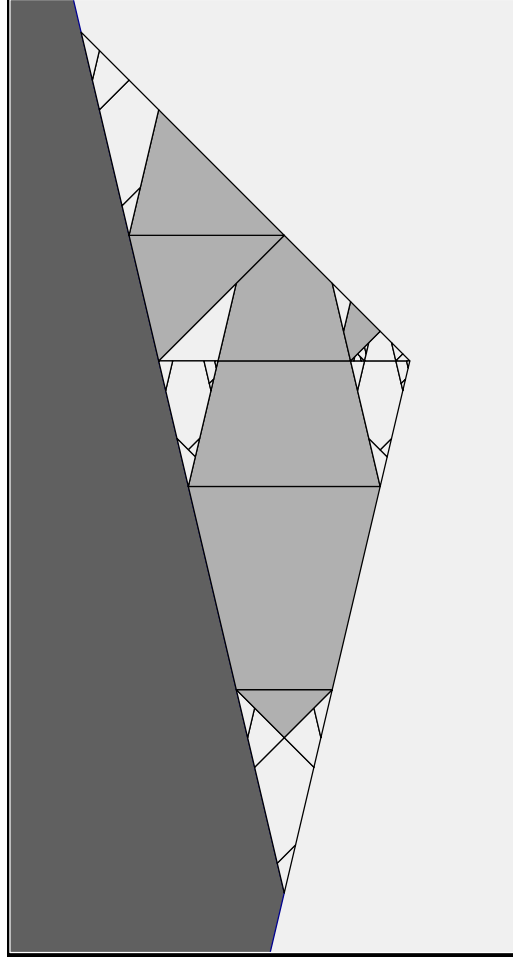


Figure 11.1: The partition of T_{11}^+ and an edge-covering.

The A -partition of T_{11}^+ is a union of $41 = 34 + 7$ A -atoms and periodic tiles. The periodic tiles are all shaded. The dark polygon on the side is part of the Penrose kite. The horizontal lines through the periodic tiles correspond to the tops and bottoms of the layers of \hat{A} . Thus, for instance, even though the large octagon is a single periodic tile, we treat it as 4 separate tiles in our proof. Each of the 3 triangles T_{12}^+ , T_{31}^+ and T_{41}^+ are also partitioned into $41 = 43 + 7$ polygons. These partitions look exactly like the one in Figure 11.1, up to reflection in a vertical line. The triangle T_{41}^+ is partitioned into $4 = 3 + 1$ polygons and T_{41}^- is partitioned into $3 = 2 + 1$ polygons. It turns out that we do not need to consider the 4 remaining triangles, as we explain below. The reader can see all the partitions, and in color, using our applet.

Lemma 11.1 *Suppose that Equation 103 holds for all the A -atoms in the A -partitions. Then the Fundamental Orbit Theorem is true.*

Proof: Let O_1 be a generic orbit that intersects T_{ij}^\pm but avoids set $\mathcal{T}(2)$. Suppose first that O_1 is not a distinguished orbit. Since all the periodic tiles in the A -partitions lie in $\mathcal{T}(2)$, some point $p_1 \in O_1$ intersects an A -tile τ of the A -partition. Since p_1 is generic, it cannot intersect the boundary of an A -atom. Hence p_1 intersects the interior of some A -atom τ of the A -partition. But then, by Equation 103, we have $O_1 \rightsquigarrow O_2$, where O_2 is the Ψ -orbit containing $p_2 = \psi^{-1}(T_{ij}^\pm(p_1))$.

When O_1 is a distinguished orbit, we attach it to one of the slabs or the other. The argument above is the same, except that what we get is that p_1 must lie in the interior of one of the horizontal edges of the A -partition. But then Equation 103 gives us the same result as for the ordinary orbits. Were we to attach the special orbit to the other slab, we could still use Equation 103, but it would involve a different A -atom. ♠

To prove the Fundamental Orbit Theorem, it only remains to check Equation 103 for each of the A -atoms in the A -partitions. Each individual check is a finite calculation, similar to what we did in the last chapter. We now discuss this calculation.

First of all, we reduce the calculation from 10 triangles to 6 triangles, as follows. We have the relations

$$\Psi(T_{11}^-) = T_{12}^+; \quad \Psi(T_{12}^-) = T_{11}^+; \quad \Psi(T_{31}^-) = T_{32}^+; \quad \Psi(T_{32}^-) = T_{31}^+. \quad (104)$$

Using these relations, we see right away that the truth of the Fundamental Orbit Theorem for each triangle on the right hand side of the relation implies the truth of the Fundamental Orbit Theorem for the triangle on the left hand side. Thus, we only need to check the 5 triangles T_{ij}^+ and the triangle T_{41}^- .

First of all, we check that each relevant triangle T_{ij}^\pm is indeed partitioned by the polygons we have listed. The check is similar to what we did in §6. As for verifying Equation 103, we have to make the same verification $141 = 4 \times 34 + 3 + 2$ times.

Let P and Q respectively be the polygons on the left and right hand sides of Equation 103. To check that $P \rightarrow Q$, in each of the 141 cases, we do the following calculation.

- We check that $\widehat{P} = \Theta(P)$ is contained in a \widehat{A} -atom. This amounts to producing the two integers m and n , with $m \leq 0 < n$, such that $\widehat{\Psi}^k$ is entirely defined on the interior of \widehat{P} for all $k \in [m, n]$, and $\widehat{\Psi}^k(\widehat{P}) \subset \widehat{A}$ for $k = m, n$ but not for $k \in (m, n)$. We verify these things using the methods discussed in §7.5. Define $\widehat{P}_0 = \widehat{\Psi}^m(\widehat{P})$.
- We check that $\widehat{Q} = \Theta(Q)$ is contained in a \widehat{B} -atom. This is the same kind of check as in the previous case. Let $\widehat{Q}_0 = \widehat{\Psi}^m(\widehat{Q})$ be the iterate that is contained in \widehat{B} . (Depending on the layer, m is either non-positive or non-negative.)
- We check that $\widehat{R}(\widehat{P}_0) = \widehat{Q}_0$.

We simply perform the 141 calculations and observe that they all work. This completes the proof of the Fundamental Orbit Theorem.

Remark: Note that we are not quite verifying that \widehat{P} and \widehat{Q} are atoms, but only that they are contained in atoms. Similarly, we are only verifying that P and Q are contained in atoms, and that their boundaries are all contained in lines defined over $\mathbf{Z}[\phi]$. This suffices for our purposes. We mention, however, that we produce our A -partitions by slicing the \widehat{A} -atoms and periodic tiles by the image of Σ . Thus, were to make the extra check, it would work out.

The reader can use our applet to survey the calculations. Here we give an example illustrating the rough size and complexity of the calculation. There is a small isosceles A -tile in the partition of T_{11}^+ that contains the rightmost vertex of T_{11}^+ . This triangle is perhaps too small to see in Figure 11.1. The vertices of this triangle are

$$(997 - 616\phi, -236 + 146\phi) \quad (1984 - 1226\phi, -3 + 2\phi) \quad (10 - 6\phi, -3 + 2\phi).$$

In this case

$$\widehat{P}_0 = \widehat{\Psi}^{-508}(\widehat{X}); \quad \widehat{Q}_0 = \widehat{\Psi}^{87}(\widehat{Q}).$$

Things don't get much worse than this, because the maximum chain length is 703.

11.3 The Fixed Point Theorem

It turns out that there is a second region that plays a role that is similar to the fundamental triangles. Let Q_0 be the quadrilateral with vertices

$$(6 - 2\phi, \phi^{-3}); \quad (3, 0); \quad (2\phi, \phi^{-3}); \quad (3, 2\phi^{-1}) \quad (105)$$

Q_0 is a small kite whose line of symmetry is the vertical line $x = 3$. The bottom point of Q is $(3, 0)$. See Figure 11.2 below. Let D denote the homothety that fixes $(3, 0)$ and expands distances by a factor of ϕ^3 . Define

$$Q_n = D^{-n}(Q). \quad (106)$$

The family $\{Q_n\}$ is a shrinking family of kites that limits to $(3, 0)$. Let ρ denote the reflection in the vertical line $x = 3$.

Theorem 11.2 (Fixed Point) *Let O_1 be any sufficiently long and generic orbit that intersects Q_1 . Then $O_1 \rightsquigarrow O_2$, and O_2 is such that*

$$\psi'(\overline{O_2}) \cap Q_0 = \rho \circ D(O_1 \cap Q_1).$$

Here O_2 is the reflection of O_2 in the x -axis and ψ' is the outer billiards map. In particular $O_1 \rightarrow O_2$, in the sense of Theorem 6.4.

Proof: The proof here is essentially the same as the proof we gave for the Fundamental Orbit Theorem. Let $\Delta : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be the dilation by ϕ^3 such that $\Delta(3, 0) = (-3, -2)$. We have the equation

$$\psi' \circ \overline{\Delta(p)} = \rho \circ D(p) \quad (107)$$

for all $p \in Q_1$.

We exhibit a tiling of Q_1 by a union of 48 polygons, 42 of which are A -atoms and 6 of which are periodic tiles. For each A -tile τ in the partition, we verify, by the same means as above, the equation

$$\tau \rightarrow \Delta(\tau), \quad (108)$$

Our result follows from Equation 107, Equation 108, and the same argument as in Lemma 11.1. ♠

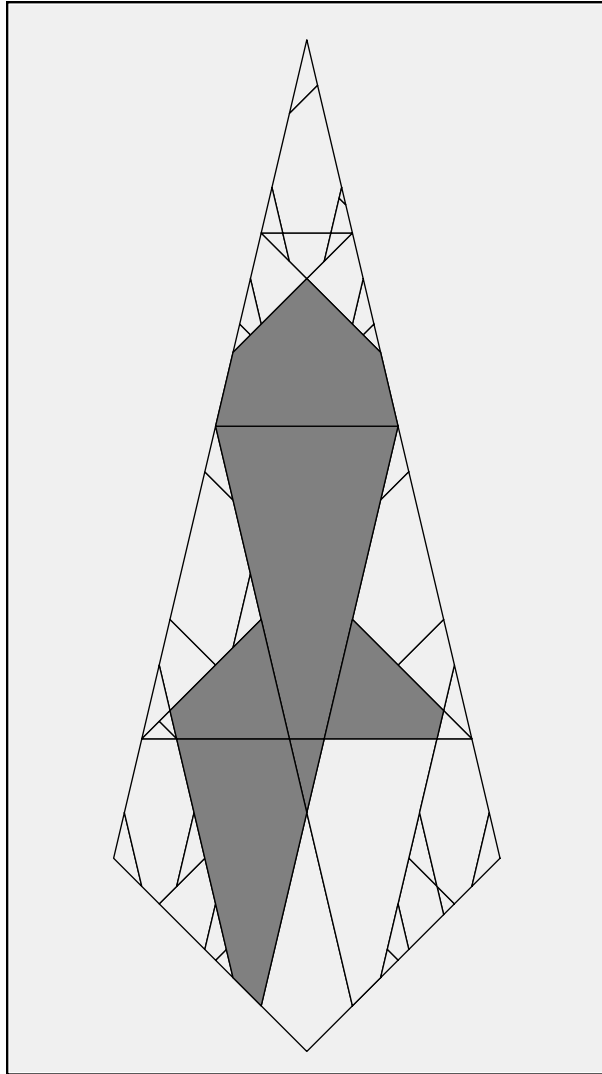


Figure 11.2: Q_1 and its partition into atoms and periodic tiles.

It is a consequence of the Fixed Point Theorem that the point $(3, 0)$ is a point of dilation symmetry for the dynamical tiling associated to outer billiards on the penrose kite. Figure 1.3 shows what is going on. The white regions in Figure 1.3 are small copies of portions the tiling \mathcal{T} and the shaded regions are increasingly small periodic tiles. The reader can probably spot Q_1 in Figure 1.3: It is flanked by the two largest dark tiles.

12 Proof of the Near Reduction Theorem

12.1 Discussion

The main structural relation between the \hat{A} -atoms and the \hat{B} -atoms is the 3-to-1 map from the set of \hat{A} -chains to the set of \hat{B} -chains. Now we mention a second piece of structure. Though we did not formally check this in general, in the cases of interest it turns out that each \hat{B} -atom is partitioned into a finite union periodic tiles and \hat{A} -atoms. Thus, we have a kind of auxilliary dynamical system defined on the level of \hat{A} -atoms. Starting with a \hat{A} -atom τ , we can consider all those \hat{B} -atoms – there are finitely many, and all in the same chain – such that $\tau \rightarrow \tau'$. Next, we subdivide each such τ' back into \hat{A} -atoms. And so on. All our proofs in this chapter amount to tracing through this auxilliary dynamical system.

Our computer program is designed so that the user can trace through all the calculations manually, plotting the relevant tiles.

The way to view this chapter is that we have one main calculation, contained in §12.4, that takes care of almost all the orbits that come fairly close to the origin. The remaining calculations, such as those in §12.3, deal with the several small regions that the big calculation does not cover.

12.2 Points Very Near the Kite

In our next result, the notation $O_1 \rightarrow O_2$ is as in Theorem 6.4. Let

$$S_r = [0, r] \times [0, 2]. \quad (109)$$

Lemma 12.1 *Let $r = 2 + \phi^{-3}$. Let O_1 be any generic infinite orbit that intersects S_r . either O_1 intersects T^+ or $O_1 \rightarrow O_2$ and O_2 intersects T^+ .*

Proof: Our proof refers to Figure 12.1. Let $r = 2 + \phi^{-3}$. Let ρ denote complex conjugation – i.e. reflection in the x -axis. Figure 12.2 shows a covering of S_r by 13 tiles. The left black tile is the Penrose kite, and the right vertex of the rightmost dark tile lies on the vertical line $y = r$. The two white tiles comprise T , the fundamental triangle. Let X_1 be the lower of the two black tiles and let L be the smallest light tile – namely, the little kite. Note that X_1 and L share an edge and $X_1 \cup L$ is a triangle.

Short calculations reveal the following.

- Ψ is the identity on all the light tiles.
- The top black tile X_2 is such that $(\psi')^{-1} \circ \Psi^2(X_2) \subset T$.
- The leftmost dark tile D_1 is such that $\Psi(D_1) \subset T$.
- The bottom dark tile D_2 is such that $\rho \circ \Psi(D_2) = X_1$.
- The rightmost dark tile D_3 is such that $\rho \circ \Psi(D_3) = X_1 \cup L$.

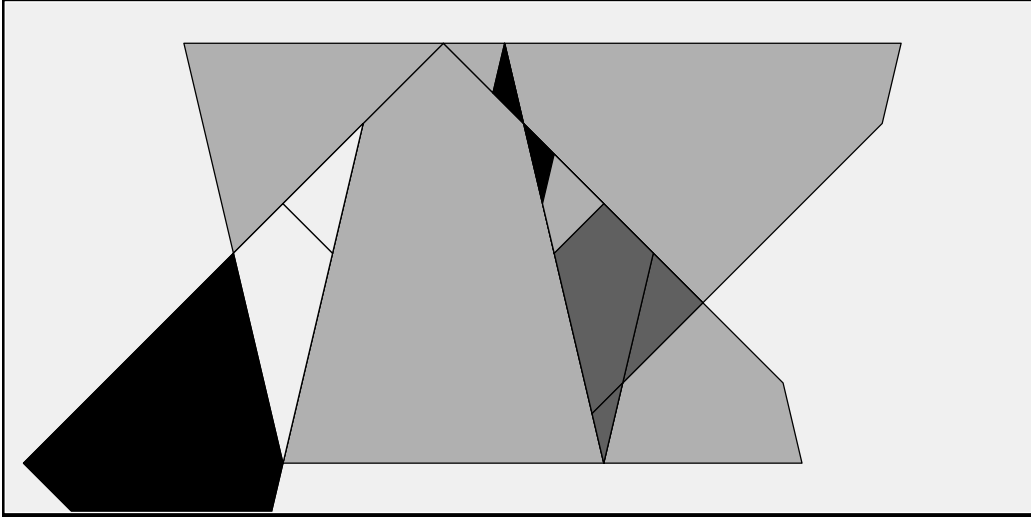


Figure 12.1: A covering of Σ_r .

The region X_1 is the triangle T_{41}^- analyzed in the proof of the Fundamental Orbit Theorem. By the Fundamental Orbit Theorem, the orbits intersecting X_1 renormalize to orbits that intersect T . Lemma 12.1 now follows directly from case-by-deductions. We analyze of the cases in detail. The remaining cases are similar, and in fact easier. (We chose the most elaborate case.) Suppose that O_1 intersects D_3 , the rightmost dark tile. Then an associate O'_1 intersects $L \cup X_1$. Being an infinite orbit, O'_1 does not intersect the interior of L . Being a well-defined orbit, O'_1 does not intersect $\partial L \cup \partial X_1$. Hence O'_1 intersects the interior of X_1 . But then $O'_1 \rightsquigarrow O_2$, where O_2 intersects T . This shows that $O_1 \rightarrow O_2$ where O_2 intersects T . ♠

12.3 Two Trouble Spots

Let σ be the triangle with vertices

$$(-15 + 11\phi, 2 - \phi) \quad (-36 + 24\phi, -19 + 12\phi) \quad (53 - 31\phi, 2 - \phi). \quad (110)$$

Let σ^* be the triangle with vertices

$$(21 - 11\phi, 2 - \phi) \quad (42 - 24\phi, -19 + 12\phi) \quad (-47 + 31\phi, 2 - \phi). \quad (111)$$

σ and τ are two small A -atoms contained in $Q_0 - Q_1$. Reflection in the vertical line $x = 3$ interchanges them. Let ψ' denote the outer billiards map.

Lemma 12.2 *Let O_1 be any generic infinite orbit that intersects σ . Then we have $O_1 \rightsquigarrow O_2 \rightsquigarrow O_3$ and $\psi'(O_3)$ intersects T^- .*

Proof: Let $\sigma_1 = \sigma$. Using the notation above, and the same kind of calculations, we check that $\sigma_1 \rightsquigarrow \sigma_2$, where σ_2 is the triangle (a translate of σ_1) with vertices

$$(-11 + 9\phi, -2 + \phi) \quad (-32 + 22\phi, -23 + 14\phi) \quad (57 - 33\phi, -2 + \phi).$$

We check that σ_1 is a union of one periodic tile and 2 A -atoms, σ_2 and σ_3 . Next, we check that $\sigma_2 \rightarrow \sigma_4$ and $\sigma_3 \rightarrow \sigma_5$, where $\sigma_6 = \sigma_4 \cup \sigma_5$ is the triangle with vertices

$$(5 - 4\phi, -6 + 4\phi) \quad (18 - 12\phi, 7 - 4\phi) \quad (-3 + \phi, 2 - \phi).$$

A direct calculation shows that $\psi'(\sigma_6) \subset T^-$. An argument just like the one given in Lemma 11.1 finishes the proof. ♠

Lemma 12.3 *Let O_1 be any generic infinite orbit that intersects σ^* . Then we have $O_1 \rightsquigarrow O_2 \rightsquigarrow O_3 \rightsquigarrow O_4$ and $\psi'(O_4)$ intersects T^+ .*

Proof: The argument here is very similar to what we did in the previous lemma. We omit the details. ♠

12.4 The Big Calculation

Now we prove a result similar to the ones in the previous section, except that the computational part of the proof is much more intensive.

Given any $p = (x, y)$, define $|p|_x = |x|$. In our next result, the significance of the number $2\phi^{-3}$ is that it is precisely the horizontal width of Q_0 .

Lemma 12.4 *Let $r = 2 + \phi^{-1}$. Let $p_1 \in S_{24} - S_r - Q_1 - \sigma - \sigma^*$ be a point with a generic infinite orbit O_1 . Then $O_1 \rightsquigarrow O_2 \rightsquigarrow O_3$, and there is some $k \in \{1, 2, 3\}$ and some $q_k \in O_k$ such that $|q_k|_x \leq |p_k|_x - 2\phi^{-3}$.*

Proof: Say that a *symmetric return tile* is a maximal polygon in the main strip Σ on which both Ψ and Ψ^{-1} are defined. For each point p in a symmetric return tile τ , the expressions

$$K_{\pm}(\tau) = |\Psi^{\pm 1}(p)|_x - |p|_x$$

are independent of the choice of p . If $f(\tau) = 0$ it means that Ψ is the identity on τ . Otherwise, by Equation 36, we have $\min |K_{\pm}(\tau)| \leq 2\phi^{-3}$. Accordingly, we only have to consider points in symmetric return tiles τ such that $K_+(\tau)$ and $K_-(\tau)$ are both positive. Call such tiles *positive symmetric return tiles*, or PSRT's for short.

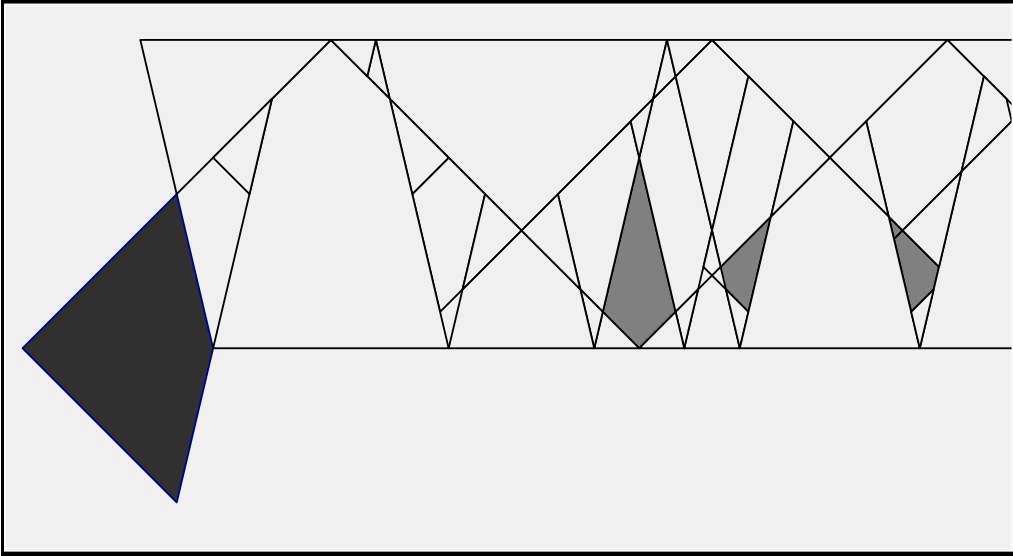


Figure 12.2: Four of the PSRT's.

It turns out that there are 23 PSRT's that intersect S_{24} . The PSRT closest to the origin is Q_0 . Figure 12.2 shows the first 4 PSRT's, shaded, as well as a portion of the partition of the top half of Σ into symmetric return tiles. (The rightmost shaded piece is a union of two tiles.) The reader can see the remaining tiles using our applet.

We partition the union of 23 PSRT's by 848 polygons, each of which is either an A -atom or a periodic tile. Since we are only interested in infinite orbits, we ignore the periodic tiles. Let τ be an A -atom in our partition, and let $[\tau]$ be the chain containing τ . Say that τ is *slack* if there is some other A -atom $\tau' \subset [\tau]$ such that

$$\max_{p' \in \tau'} |p'|_x \leq \max_{p \in \tau} |p|_x - 2\phi^{-3}. \quad (112)$$

Otherwise, we call τ *taut*. A taut atom is one that is nearly as close as possible to the origin within its chain. Note that τ' need not be in our partition.

Ignoring the atoms in Q_1 , which we do not need to consider for this lemma, we find that there are 123 taut atoms. With 12 exceptions, we take each taut atom τ and exhibit a B -atom τ' such that $\tau \rightarrow \tau'$ and the pair (τ, τ') satisfies Equation 112. Call the 12 exceptional atoms *super taut*. The 12 super taut atoms cannot be brought substantially closer to the origin by a single renormalization.

To deal with the 12 super taut atoms, we trace through the second renormalization, as follows. Let τ be a super taut atom.

- We find a B -atom τ' such that $\tau \rightarrow \tau'$. In practice, we take τ' to be the atom in its chain which is closest to the origin.
- The atom τ' is partitioned into a finite union τ'_1, \dots, τ'_k of A -atoms and periodic tiles. The value k depends on τ . The largest value is 27.
- For each periodic tile τ'_k we exhibit a B -atom τ''_i such that $\tau'_i \rightarrow \tau''_i$ and (τ''_i, τ) satisfies Equation 112. Again, we take τ''_k so that it minimizes the distance to the origin within its chain.

The analysis above handles the points in 10 of the 12 super-taut atoms.

The two exceptional super taut atoms are σ and σ^* , which we have deliberately excluded in our hypothesis. ♠

12.5 Proof of the Near Reduction Theorem

The Near Reduction Theorem concerns generic infinite orbits that intersect the rectangle $\Sigma_{24} = [-24, 24] \times [-2, 2]$, but our results above concern the smaller region $S_{24} = [0, 24] \times [0, 2]$. However, replacing an orbit O_1 by one of its 3 associate orbits, we can assume that the orbit intersects S_{24} . Let O_1 be such an orbit.

Since O_1 is a generic infinite orbit, we have $O_1 \rightsquigarrow O'_2 \rightsquigarrow O'_3 \dots$. These orbits intersect the strip $\Sigma = \mathbf{R} \times [-2, 2]$ but they might not intersect the smaller region $S = \mathbf{R}^+ \times [0, 2]$. However, replacing our orbits with associates, if necessary, we get a new sequence $O_1 \rightarrow O_2 \rightarrow O_3 \dots$ in which each orbit intersects S .

Remark: We allow ourselves the liberty of passing to associates mainly for convenience. We might have proved the Near Reduction Theorem for the original sequences of renormalizations, but it is more computational work.

Say that an orbit O is *passed* Q_1 if there is some $p \in O$ such that

$$|p|_x \leq \min_{q \in Q_1} |q|_x.$$

Lemma 12.5 *The Near Reduction Theorem holds for a generic orbit O_1 that is passed Q_1 .*

Proof: Iterating Lemma 12.4, we find that there is some m such that one of three things is true.

1. O_m intersects S_r for $r = 2 + \phi^{-3}$.
2. O_m intersects $\sigma \cup \sigma^*$, the union of two special A -atoms.

In either case, the lemmas in the preceding section combine to show that O_{m+k} intersects T^+ for some $k \in \{0, 1, 2, 3\}$. ♠

Now let's consider the general case when O_1 is not necessarily past Q_1 . We can still iterate Lemma 12.4, but now we arrive at the possibility that O_m intersects Q_1 for some m . The map Ψ is the identity on the line $y = 0$, so O_m does not intersect this line. Hence, there is some $\epsilon > 0$ such that O_m avoids the ϵ neighborhood of the point $(3, 0)$. But then there is some n such

that O_m intersects Q_n but not Q_{n+1} . Applying the Fixed Point Theorem, we see that O_{m+1} intersects Q_{n-1} but not Q_n , and O_{m+2} intersects Q_{n-2} but not Q_{n-1} . Continuing in this way, we see that $O_1 \rightarrow \dots \rightarrow O_m \rightarrow \dots \rightarrow O_h$, where O_h intersects Q_0 but not Q_1 .

If O_h intersects $S_r \cup \sigma \cup \sigma^*$ then we have the same situation as in Lemma 12.5. Otherwise, we can apply Lemma 12.4 one last time to produce an orbit O_{h+k} which is passed Q_1 . This works thanks to the following three facts.

- O_h intersects Q_0 in an interior point (because the orbit is well-defined).
- $|O_{h+k}|_x \leq |O_h|_x - 2\phi^{-3}$.
- Q_0 has width $2\phi^{-3}$.

We finish the proof by applying Lemma 12.5 to O_{h+k} .

For each orbit O_1 , the Near Reduction Theorem produces some integer $m = m(O_1)$ such that $O_1 \rightarrow \dots \rightarrow O_m$, and O_m intersects T^+ . There is no uniform m which works for all orbits. The difficulty is that orbits that come sufficiently near $(3, 0)$ do not intersect T^+ , and iterated renormalization only gradually moves such orbits away from $(3, 0)$, as a consequence of the Fixed Point Theorem. On the other hand, our analysis above yields the second statement of the Near Reduction Theorem:

Lemma 12.6 *Let $y \in (0, 2)$ be any point such that $\{R^n(y)\}$ does not contain 0. Then there is some N with the following property. If O_1 contains a point within $1/N$ of the segment $[-24, 24] \times \{y\}$, then $O_1 \rightarrow \dots \rightarrow O_m$ and O_m intersects T^+ and $m < N$.*

Proof: Let's consider iteratively applying Lemma 12.4 to O_1 . After we apply Lemma 12.4 at most $24/\phi^{-3}$ times, the resulting orbit is either passed Q_1 or else intersects Q_1 . In the latter case, there is some ϵ such that every point of the resulting orbit is at least ϵ from $(3, 0)$. The number ϵ works uniformly as long as O_1 contains a point sufficiently close to our line segment. But then, an additional $C \log(\epsilon)$ renormalizations produce an orbit that intersects T^+ . Here C is some uniform constant. ♠

13 Proof of the Far Reduction Theorem

13.1 Deviation Estimates

Let $\Theta : \Sigma \rightarrow \widehat{\Sigma}$ be as in the Compactification Theorem. Recall that an A -core is a finite portion $\alpha = \{p_0, \dots, p_n\}$ of a Ψ -orbit with the following properties.

1. $\Theta(p_0) \in \widehat{A}$, the first of the renormalization sets.
2. $\Theta(p_1), \dots, \Theta(p_n)$ do not lie in \widehat{A} .
3. $\Theta(p_{n+1}) \in \widehat{A}$.

Up to translation, there are only finitely many A -cores, one per \widehat{A} -chain. To the \widehat{A} -core α we associate the *deviation interval*

$$\langle \alpha \rangle = \left[\min_k \pi_1(p_k - p_0), \max_k \pi_1(p_k - p_0) \right]. \quad (113)$$

Here π_1 is projection onto the first coordinate. That is, we translate the A -core so that its initial point is at the origin, and we measure how far the sequence deviates to the left and right of the origin.

The sequence of points $\{p_k - p_0\}$ associated to α is combinatorial in nature: In view of Equation 36, there is a finite sequence $\{(a_k, b_k)\}$ such that

$$\pi_1(p_k - p_0) = 2Aa_n + 2b_k; \quad k = 1, \dots, n. \quad (114)$$

Indeed, we can form a connected lattice path by connecting the consecutive integers $(0, 0) = (a_0, b_0), \dots, (a_n, b_n)$. We call this path an A -strand. An A -strand is nothing more than an arc of the arithmetic graph, translated to that the initial point is the origin. The A -strands are not essential for our analysis, but we introduce them because they help us visualize what is going on.

For the purposes of drawing pictures, we will continue the strands to the point (a_{n+1}, b_{n+1}) , so that the endpoints of the strand correspond to points in \widehat{A} . In this way, the arithmetic graph component corresponding to an infinite orbit is partitioned into uniformly short arcs, with the endpoints being special points that map into \widehat{A} . The distribution of these special *marker points* controls the coarse geometry of the arithmetic graph, and thereby the coarse geometry of the orbit itself.

We define B -cores and B -strands in the same way, except that in the type-2 case, the last point of the B -core maps into \widehat{B} . The Renormalization Theorem gives us a 3-to-1- map from the set of A -cores to the set of B -cores. As in the Far Reduction Theorem, the notation $\alpha \rightsquigarrow \beta$ indicates that the B -core β is the image of the A -core α under this map.

Let $N_k(S)$ denote the k -tubular neighborhood of a set S . Let $\phi^{-3}I$ denote the interval obtained by scaling I down by a factor of ϕ^{-3} about the origin.

Lemma 13.1 (Trivial) *There is a universal constant U with the following property. Suppose that $\alpha \rightarrow \beta$.*

1. *If α has type 1 then $\langle \alpha \rangle \subset N_U(\langle \beta \rangle)$.*
2. *If α has type 2 then $\phi^{-3}\langle \alpha \rangle \subset N_U(\langle \beta \rangle)$.*
3. *If α has type 1 then $\langle \beta \rangle \subset N_U(\langle \alpha \rangle)$.*
4. *If α has type 2 then $\langle \beta \rangle \subset N_U(\phi^{-3}\langle \alpha \rangle)$.*

Proof: Up to translation there are only finitely many cores. Hence, all deviation intervals are contained in a single compact subset of \mathbf{R} . ♠

Remark: There is nothing special about ϕ^{-3} in the Trivial Lemma. Were we to pick any other constant, we would get a similar result. However, we place ϕ^{-3} in this result for the sake of comparison with the more effective result that we establish below. For the more effective result, the choice of ϕ^{-3} is the natural choice.

Items 3 and 4 serve our needs perfectly. They correspond to the constant C in the Far Reduction Theorem that we don't care about. However, we do need a more effective version of Items 1 and 2.

Lemma 13.2 (Deviation) *Suppose that $\alpha \rightarrow \beta$.*

1. *If α has type 1 then $\langle \alpha \rangle \subset N_r(\langle \beta \rangle)$. Here $r = -20 + 24\phi < 12 - 2\phi$.*
2. *If α has type 2 then $\phi^{-3}\langle \alpha \rangle \subset N_5(\langle \beta \rangle)$.*

Proof: We simply enumerate all the integer sequences and check that the bounds hold. We use golden arithmetic to make sure that everything is computed correctly. ♠

Remark: We could get similar improvements on Items 3 and 4 of the Trivial Lemma, but we don't care about those cases.

13.2 An Estimate for the Markers

The Trivial Lemma and the Deviation Lemma reduce the Far Reduction Theorem to a statement that just concerns points that map into \widehat{A} and \widehat{B} , as we now explain. Say that an *A-marker* is a point $p \in \Sigma$ such that $\Theta(p) \in \widehat{A}$. We define the type of the point just as we defined the type of an *A-core*: It depends on which layer of \widehat{A} contains the image.

We write $a \rightsquigarrow b$ if a is an *A-marker* and b is a *B-marker*, and

$$\widehat{R} \circ \Theta(a) = b. \quad (115)$$

By definition $a \rightsquigarrow b$ if and only if $\alpha \rightsquigarrow \beta$, where α is the *A-chain* containing a and β is the *B-chain* containing b . Below we will prove the following result.

Lemma 13.3 (Marker) *Suppose that $a \rightsquigarrow b$.*

1. *If a has type 1 then $|b|_x \leq |a|_x + 2\phi$.*
2. *If a has type 2 then $|b|_x < \phi^{-3}|a|_x + 10$.*

The lower bounds in the Far Reduction Theorem are immediate consequences of the triangle inequality, the Marker Lemma, and Items 3 and 4 of the Deviation Lemma. Let's establish the upper bound $|\beta|_x < |\alpha|_x + 10$ in the type-1 case. Let a and b respectively be the markers in α and β . Choose the point $a' \in \alpha$ that realizes $|\alpha|_x$. Set $r = 12 - 2\phi$. We have

$$|a'|_x - |a|_x \in \langle \alpha \rangle \subset N_r(\langle \beta \rangle). \quad (116)$$

Therefore, there is some $b' \in \beta$ such that

$$|b'|_x - |b|_x \leq |a'|_x - |a|_x + (12 - 2\phi). \quad (117)$$

By the Marker Lemma,

$$|b'|_x - |a|_x < |b'|_x - |b|_x + 2\phi. \quad (118)$$

Therefore

$$|b'|_x - |a|_x < |a'|_x - |a|_x + 12. \quad (119)$$

Adding $|a|_x$ to both sides, we get

$$|\beta|_x \leq |b'|_x < |a'|_x + 12 = |\alpha|_x + 12, \quad (120)$$

as desired. The proof of the upper bound in the type-2 case works the same way.

The remainder of the chapter is devoted to proving the Marker Lemma. The type-1 case turns out to be fairly trivial, and the type-2 case has an arithmetical feel to it.

13.3 The Type-1 Case

Let us recall the information given in §5. First of all, we have

$$\Theta(x, y) = \left(1, \frac{1}{2}, 0\right) + \left(\frac{x}{\phi}, \frac{x-y}{2}, y\right) \quad (121)$$

The image of Θ is a certain fundamental domain of the form $F \times [0, 2]$, where F is described in §5.2. For our purposes, we can deal with the simpler region

$$\Omega = [1/2, 3/2] \times [0, 2] \times [0, 2] \subset F \times [0, 2] \quad (122)$$

because (by direct inspection) this region contains both \widehat{B} and \widehat{A} .

In the type 1 case, namely in the slabs corresponding to the partition intervals I_2 and I_4 , the formula for \widehat{R} is

$$\widehat{R}(x, y, z) = (x, y, z \pm 2\phi^{-1}). \quad (123)$$

In this case, we can restrict \widehat{R} to a smaller domain $S \subset \Omega$ such that $\widehat{R}(S) \subset \Omega$ as well.

Given the explicit equation for \widehat{R} , we get the following result. If $a = (x, y)$ and $a \rightarrow b$, then

$$b = (x \mp 2\phi, y \pm 2\phi^{-1} \pm 2). \quad (124)$$

What makes this equation work is that $2\phi - 2\phi^{-1} \equiv 0 \pmod{2\mathbf{Z}}$. Once we see that this formula works, it must be the uniquely correct formula, because Θ is an injective map.

Equation 124 immediatly implies the type-1 case of the Marker Lemma.

13.4 The Type-2 Case

The type-2 case involves the 7 maps \widehat{R}_{ijk} , for various choices of i, j, k . These maps are listed in §5.4. Define

$$S_{ijk} = \widehat{R}_{ijk}^{-1}(\Omega). \quad (125)$$

Here S_{ijk} is a small rectangular solid that either contains or abuts the fixed point (i, j, k) of \widehat{R}_{ijk} . In the Renormalization Theorem, the map \widehat{R}_{ijk} is only applied to polyhedra that are entirely contained in S_{ijk} .

Since \widehat{R}_{ijk} expands distances by a factor of ϕ^3 , we see that the sidelength of S_{ijk} in the x -direction is ϕ^{-3} and the sidelengths in the other directions are $2\phi^{-3}$.

For the sake of notational convenience, we will give our proof in the case of \widehat{R}_{111} . The other cases have essentially the same proof.

Let $p = (p_1, p_2)$ be some point. Let $\Theta_k(p)$ denote the k th coordinate of $\Theta(p)$. We first observe that $\Theta_1(p) = 1$ if and only if $p_1 = 2N\phi$ for some integer N . So, we can always write

$$p_1 = 2N_p\phi + \epsilon_p, \quad (126)$$

where ϵ_p is an error term of size at most ϕ .

When $p = a$, we have a better estimate. Since $\Theta_1(a)$ lies within $\phi^{-3}/2$ of 1, the error term ϵ_a is at most $\phi^{-2}/2$. We also observe that a_2 either lies within ϕ^{-3} of 1 or within ϕ^{-3} of -1 . We will treat the former case. The latter case has the same treatment. Summarizing the discussion, we have

$$a = (2N_a\phi + \epsilon_a, 1 + \delta_a); \quad |\epsilon_a| \leq \phi^{-2}/2; \quad |\delta_a| \leq \phi^{-3}. \quad (127)$$

From the equation

$$\Theta_k(b) - 1 = \left(\Theta_k(a) - 1 \right) \times \phi^3, \quad (128)$$

we get

$$b = (2N_b\phi + \phi^3\epsilon_a, 1 + \phi^3\delta_a). \quad (129)$$

In the formula for Θ , we take the listed coordinates and suitably translate them by even integers until these coordinates all lie in $[0, 2]$. Hence, there is an even integer M'_a such that

$$\Theta_2(a) = N_a\phi + M'_a + \epsilon_a/2 - \delta_a/2 \in [0, 2].$$

We write $M'_a = 1 + M_a$, so that

$$\Theta_2(a) = 1 + N_a\phi + M_a + \epsilon_a/2 - \delta_a/2 \quad (130)$$

Similarly,

$$\Theta_2(b) = 1 + N_b\phi + M_b + \phi^3(\epsilon_b/2 - \delta_b/2). \quad (131)$$

Given our bounds on ϵ_a and δ_a , we have

$$M_a \in [-N_a\phi - 2, -N_a\phi + 2]. \quad (132)$$

Combining the $k = 2$ case of Equation 128 with Equations 130 and 131, we get

$$N_b\phi + M_b = \phi^3(N_a\phi + M_a). \quad (133)$$

Expanding out the right hand side of this last equation, we get

$$\begin{aligned} N_b\phi + M_b &= \phi^4 N_a + \phi^3 M_a = \\ &= (2 + 3\phi)N_a + (1 + 2\phi)M_b = \\ &= (3N_a + 2M_b)\phi + (2N_a + N_b). \end{aligned} \quad (134)$$

Equating coefficients, we get

$$N_b = 3N_a + 2M_a. \quad (135)$$

Combining this last equation with Equation 132, we find that

$$N_b \in [(3 - 2\phi)N_a - 2, (3 - 2\phi)N_a + 2] = [\phi^{-3}N_a - 2, \phi^{-3}N_a + 2].$$

In short,

$$\phi^{-3}|N_a| \leq |N_b| + 2. \quad (136)$$

Multiplying through by 2ϕ , we get

$$\phi^{-3} \times (2\phi|N_a|) \leq 2\phi|N_b| + 4\phi. \quad (137)$$

Finally, we mention that

$$\left| |a|_x - 2\phi|N_a| \right| = \epsilon_a \leq \phi^{-2}/2; \quad \left| |b|_x - 2\phi|N_b| \right| = \epsilon_b \leq \phi/2; \quad (138)$$

By the triangle inequality, we get

$$b_x \leq \phi^{-3}a_x + 4\phi + \phi/2 + \phi^{-3} \times (\phi^{-2}/2) < 10. \quad (139)$$

This completes the proof of the type-2 case of the Marker Lemma.

14 Coordinates

14.1 The Polyhedron Exchange Map

14.1.1 Notation and Conventions

The polyhedron exchange map is defined in terms of 64 polyhedra P_0, \dots, P_{63} . The polyhedra P_0, \dots, P_{21} are what we call *inactive*, in the sense that $\widehat{\Psi}$ acts as the identity on these polyhedra. The remaining polyhedra are what we call *active*. The 42 active polyhedra are such that

$$P_{j+21} = I(P_j); \quad j = 22, \dots, 42. \quad (140)$$

Here I is the involution

$$I(x, y, z) = (2, 2, 2) - (x, y, z). \quad (141)$$

The map I commutes with $\widehat{\Psi}$. Thus Ψ has the same action both the lefthand and the righthand polyhedra listed in Equation 140. Accordingly, to save space, we will just list P_1, \dots, P_{42} .

Our notation is such that

$$\begin{bmatrix} a_0 & a_1 \\ b_0 & b_1 \\ c_0 & c_1 \end{bmatrix} = (a_0 + a_1\phi, b_0 + b_1\phi, c_0 + c_1\phi). \quad (142)$$

We list each polyhedron by its vertices. The vertices are not given in any special order.

14.1.2 The Inactive Polyhedra

$$P_0 = \begin{bmatrix} 3 & -1 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -3 & 3 \\ 2 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \\ \begin{bmatrix} -3 & 3 \\ 2 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P_1 = \begin{bmatrix} -1 & 1 \\ 2 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \\ \begin{bmatrix} 5 & -3 \\ 0 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P10 = \begin{bmatrix} 6 & -3 \\ 0 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ -3 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$P_{11} = \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 0 & 0 \end{bmatrix}$$

$$P_{12} = \begin{bmatrix} 11 & -6 \\ 0 & 0 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ -3 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P13 = \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P14 = \begin{bmatrix} 5 & -2 \\ 0 & 0 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -3 & 3 \\ 2 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 5 & -2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P15 = \begin{bmatrix} -3 & 2 \\ 5 & -2 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -3 & 2 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P16 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ -3 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ 3 & -1 \end{bmatrix}$$

$$P17 = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 3 & -1 \end{bmatrix}$$

$$P18 = \begin{bmatrix} 6 & -3 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ -3 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ -3 & 2 \\ 5 & -2 \end{bmatrix}$$

$$P19 = \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 2 & 0 \\ 5 & -2 \end{bmatrix}$$

$$P_{20} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -3 & 3 \\ 2 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 5 & -2 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 5 & -2 \\ -3 & 2 \\ 5 & -2 \end{bmatrix}$$

$$P_{21} = \begin{bmatrix} 2 & -1 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ 0 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -3 & 2 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -3 & 2 \\ 2 & 0 \\ 5 & -2 \end{bmatrix}$$

14.1.3 Half the Active Polyhedra

$$P22 = \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 0 & 0 \end{bmatrix}$$

$$P23 = \begin{bmatrix} 5 & -3 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$P24 = \begin{bmatrix} 3 & -1 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -5 & 4 \\ 0 & 0 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -5 & 4 \\ 5 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P25 = \begin{bmatrix} -1 & 1 \\ 0 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & -1 \\ 0 & 0 \end{bmatrix}$$

$$P26 = \begin{bmatrix} 3 & -1 \\ 0 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ -3 & 2 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ -3 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P27 = \begin{bmatrix} 8 & -5 \\ 0 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 8 & -5 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$P28 = \begin{bmatrix} 5 & -2 \\ -3 & 2 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} 5 & -2 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 3 & -1 \end{bmatrix}$$

$$P29 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ 3 & -1 \end{bmatrix}$$

$$P30 = \begin{bmatrix} -9 & 6 \\ 7 & -4 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 0 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 0 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 2 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 4 & -2 \\ 0 & 0 \end{bmatrix}$$

$$P37 = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 3 & -1 \end{bmatrix} \\ \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P38 = \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 1 & 0 \end{bmatrix}$$

$$P39 = \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 1 & 0 \end{bmatrix}$$

$$P40 = \begin{bmatrix} -11 & 8 \\ 2 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -6 & 5 \\ 2 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -6 & 5 \\ 3 & -1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -3 & 3 \\ 2 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -11 & 8 \\ 4 & -2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -3 & 3 \\ 2 & 0 \\ 2 & -1 \end{bmatrix}$$

$$P41 = \begin{bmatrix} -1 & 2 \\ 2 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -6 & 5 \\ 2 & 0 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -6 & 5 \\ 3 & -1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -11 & 8 \\ 4 & -2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -6 & 5 \\ 3 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -11 & 8 \\ 2 & 0 \\ 4 & -2 \end{bmatrix} \\ \begin{bmatrix} -6 & 5 \\ 2 & 0 \\ 4 & -2 \end{bmatrix}$$

$$P42 = \begin{bmatrix} -9 & 6 \\ 5 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 4 & -2 \end{bmatrix} \\ \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 4 & -2 \end{bmatrix}$$

Using our computer program, the reader can survey all these polyhedra. The program is set up so that the vertices of the polyhedron are displayed whenever the polyhedron is selected.

14.2 The Map

Now we describe the action of the map $\widehat{\Psi}$ on the polyhedron P_j . To each integer j we assign a pair of integers N_j according to the following lookup table.

$$\begin{bmatrix} 0 - 21 & (0, 0) \\ 22 - 29 & (1, 0) \\ 30 - 32 & (1, 1) \\ 33 - 40 & (0, 1) \\ 41 - 42 & (-1, 1) \\ 43 - 50 & (-1, 0) \\ 51 - 53 & (-1, -1) \\ 54 - 61 & (0, -1) \\ 62 - 63 & (1, -1) \end{bmatrix} \quad (143)$$

For instance, if $j = 41$ then $N_j = (-1, 1)$. The numbers $N_j = (m, n)$ are such that

$$\Psi(p) - p = (2m + 2n\phi^{-3}, *) \quad (144)$$

for any point p such that $\Theta(p) \in P_j$. As in Equation 36, the third coordinate, which lies in $\{-2, 0, 2\}$ depends on the parity of $m + n$. When $m + n = 0$ the third coordinate in Equation 144 is 0. Otherwise, it is ± 2 , depending on which value yields a point in the strip Σ .

The action of $\widehat{\Psi}$ on P_j is expressed in terms of the pair (m, n) as follows.

$$\widehat{\Psi}(\widehat{p}) - \widehat{p} = \begin{bmatrix} -2m + 10n & 2m - 6n \\ -2n & 2n \\ 0 & 0 \end{bmatrix} \quad (145)$$

Thus, for instance, for N_{41} , the pair $(-1, 1)$ yields the vector

$$(12 - 8\phi, -2 + 2\phi, 0).$$

Notice that $N_{21+j} = -N_j$ for $j = 22, \dots, 41$. This is consistent with the action of the involution I defined in the previous section.

Now we discuss the allowable transitions for our map. We write $a \rightarrow b$ is there exists a point p in the interior of P_a such that $\widehat{\Psi}(p) \in P_b$. For $a < 22$ we only have $a \rightarrow a$. We also have the general symmetry

$$a \rightarrow b \quad \Longleftrightarrow \quad (a + 21) \rightarrow (b + 21) \quad (146)$$

which holds for all $a \geq 22$. For this reason, we just list the transitions for $a = 22, \dots, 42$.

$$\begin{array}{rcl}
22 & \rightarrow & 61 \quad 23 \quad 56 \\
23 & \rightarrow & 63 \quad 58 \quad 29 \quad 32 \quad 26 \quad 24 \\
24 & \rightarrow & 25 \quad 53 \quad 37 \\
25 & \rightarrow & 31 \quad 35 \quad 28 \quad 27 \quad 23 \quad 62 \quad 61 \\
26 & \rightarrow & 25 \\
27 & \rightarrow & 63 \\
28 & \rightarrow & 25 \\
29 & \rightarrow & 23 \\
30 & \rightarrow & 23 \\
31 & \rightarrow & 32 \\
32 & \rightarrow & 25 \quad 30 \quad 47 \quad 36 \quad 34 \quad 37 \quad 42 \quad 38 \quad 39 \\
33 & \rightarrow & 32 \\
34 & \rightarrow & 46 \quad 26 \quad 32 \\
35 & \rightarrow & 44 \quad 40 \quad 41 \quad 48 \\
36 & \rightarrow & 39 \\
37 & \rightarrow & 32 \\
38 & \rightarrow & 32 \quad 33 \\
39 & \rightarrow & 32 \\
40 & \rightarrow & 48 \quad 23 \\
41 & \rightarrow & 43 \quad 32 \quad 46 \\
42 & \rightarrow & 44 \quad 40 \quad 41
\end{array} \tag{147}$$

Lemma 14.1 *Every nontrivial generic orbit intersects P_j for one of the indices $j = 23, 25, 32, 40, 41, 44, 46, 53, 61, 62$.*

Proof: Our list of indices is invariant under the involution $i \rightarrow i + 21$, which corresponds to the involution of the system, which we have mentioned several tiles. For that reason, it suffices to solve the following modified problem. We take each number $b > 42$ on the above list and replace it by $b - 21$. We then show for every $a \in [22, 42]$ that $a = a_0 \rightarrow \dots \rightarrow a_k$, such that $a_k \in L = \{23, 25, 32, 40, 41\}$. This is equivalent to the original lemma.

Here is the modified list.

$$\begin{array}{rcll}
22 & \rightarrow & 40 & 23 & 35 \\
23 & \rightarrow & 42 & 37 & 29 & 32 & 26 & 24 \\
24 & \rightarrow & 25 & 32 & 37 \\
25 & \rightarrow & 31 & 35 & 28 & 27 & 23 & 41 & 40 \\
26 & \rightarrow & 25 \\
27 & \rightarrow & 42 \\
28 & \rightarrow & 25 \\
29 & \rightarrow & 23 \\
30 & \rightarrow & 23 \\
31 & \rightarrow & 32 \\
32 & \rightarrow & 25 & 30 & 26 & 36 & 34 & 37 & 42 & 38 & 39 \\
33 & \rightarrow & 32 \\
34 & \rightarrow & 25 & 26 & 32 \\
35 & \rightarrow & 23 & 40 & 41 & 27 \\
36 & \rightarrow & 39 \\
37 & \rightarrow & 32 \\
38 & \rightarrow & 32 & 33 \\
39 & \rightarrow & 32 \\
40 & \rightarrow & 27 & 23 \\
41 & \rightarrow & 22 & 32 & 25 \\
42 & \rightarrow & 23 & 40 & 41
\end{array} \tag{148}$$

Note that $a \rightarrow b \in L$ when a is one of 26, 28, 29, 30, 31, 33, 37, 39, 42. But $27 \rightarrow 42$ and $36 \rightarrow 39$. So, we can eliminate these as well. Erasing the eliminated cases, and also erasing the cases corresponding to elements of L , we have

$$\begin{array}{rcll}
22 & \rightarrow & 40 & 23 & 35 \\
24 & \rightarrow & 25 & 32 & 37 \\
34 & \rightarrow & 25 & 26 & 32 \\
35 & \rightarrow & 23 & 40 & 41 & 27 \\
38 & \rightarrow & 32 & 33
\end{array} \tag{149}$$

Note that $34 \rightarrow 26$ or $36 \rightarrow b \in L$. Since we have eliminated 26, we can eliminate 34. Since $35 \rightarrow 27$ or $35 \rightarrow b \in L$, and we have eliminated 27, we can eliminate 35. Since $22 \rightarrow 35$ or $22 \rightarrow b \in L$, we eliminate 22. Since $24 \rightarrow 37$ or $24 \rightarrow b \in L$, and we have eliminated 37, we eliminate 24. Since $38 \rightarrow 33$ or $38 \rightarrow b \in L$, and we have eliminated 33, we eliminate 38. ♠

14.3 The set \widehat{B}

The set \widehat{B} is divided into 6 layers. These layers correspond to 6 intervals J'_1, \dots, J'_6 , which are just re-indexed versions of the levels corresponding to \widehat{B} . That is

$$J'_1 = J_1; \quad J'_2 = J_2; \quad J'_3 = J_{31}; \quad J'_4 = J_{32}; \quad J'_5 = J_4; \quad J'_6 = J_5. \quad (150)$$

The intervals J'_1, \dots, J'_6 are defined by the partition

$$0 < \phi^{-2} < 2\phi^{-2} < 1 < 2 - \phi^{-2} < 2 - 2\phi^{-2} < 2. \quad (151)$$

Each layer is decomposed into 4 convex polyhedra, which we call branches. The first 3 layers are contained in the bottom half of $\widehat{\Sigma}$, namely the region

$$(\mathbf{R}/2\mathbf{Z}) \times (\mathbf{R}/2\mathbf{Z}) \times [0, 1]. \quad (152)$$

The remaining 3 layers are the images of the first 3 layers under the involution I discussed above. For this reason, we just list the 12 polyhedra in the first 3 layers. Our notation is such that P_{ij} is the j th branch of the i th layer.

14.3.1 Layer 1

$$B_{11} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -7 & 5 \\ 6 & -3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -2 & 2 \\ 5 & -3 \end{bmatrix} \begin{bmatrix} -7 & 5 \\ 4 & -2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -7 & 5 \\ 6 & -3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 1 & 0 \\ 2 & -1 \end{bmatrix}$$

$$B_{12} = \begin{bmatrix} 9 & -5 \\ -4 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 9 & -5 \\ -2 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 0 & 0 \end{bmatrix}$$

$$B_{13} = \begin{bmatrix} 9 & -5 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 22 & -13 \\ -3 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 5 & -3 \end{bmatrix} \begin{bmatrix} 9 & -5 \\ -10 & 7 \\ 10 & -6 \end{bmatrix} \begin{bmatrix} -25 & 16 \\ 11 & -6 \\ 10 & -6 \end{bmatrix} \begin{bmatrix} -25 & 16 \\ 11 & -6 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 3 & -1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -12 & 8 \\ 3 & -1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 22 & -13 \\ -5 & 4 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 9 & -5 \\ 0 & 1 \\ 2 & -1 \end{bmatrix}$$

$$B_{14} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 9 & -5 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -12 & 8 \\ 5 & -2 \\ 2 & -1 \end{bmatrix}$$

14.3.2 Layer 2

$$B_{21} = \begin{bmatrix} -7 & 5 \\ 4 & -2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -7 & 5 \\ 6 & -3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 9 & -5 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 7 & -4 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ -1 & 1 \end{bmatrix} \\ \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 4 & -2 \end{bmatrix}$$

$$B_{22} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 9 & -5 \\ -4 & 3 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 4 & -2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 9 & -5 \\ -4 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -4 & 3 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \\ \begin{bmatrix} 9 & -5 \\ -2 & 2 \\ 2 & -1 \end{bmatrix}$$

$$B_{23} = \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 4 & -2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 7 & -4 \end{bmatrix}$$

$$B_{24} = \begin{bmatrix} -10 & 7 \\ 4 & -2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 4 & -2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 7 & -4 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 9 & -5 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 1 & 0 \\ 2 & -1 \end{bmatrix} \\ \begin{bmatrix} -10 & 7 \\ 6 & -3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 6 & -3 \\ 2 & -1 \end{bmatrix}$$

14.3.3 Layer 3

$$B_{31} = \begin{bmatrix} 6 & -3 \\ -2 & 2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 14 & -8 \\ -4 & 3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

$$B_{32} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 22 & -13 \\ -12 & 8 \\ 9 & -5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 22 & -13 \\ -7 & 5 \\ 4 & -2 \end{bmatrix}$$

$$B_{33} = \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ 9 & -5 \end{bmatrix} \begin{bmatrix} -7 & 5 \\ 4 & -2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -7 & 5 \\ 4 & -2 \\ 1 & 0 \end{bmatrix}$$

$$B_{34} = \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ -2 & 2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -15 & 10 \\ 4 & -2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 11 & -6 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 4 & -2 \\ 1 & 0 \end{bmatrix}$$

14.4 The set \hat{A} and the Renormalization Map

The set \hat{A} is divided into 18 layers, 9 of which are contained in the bottom half of $\hat{\Sigma}$. The involution I maps the first 9 layers to the second 9. For this reason, we will just explain the first 9 layers. Again, each layer has 4 branches, and so we have 36 polyhedra in all.

Each layer of \hat{A} corresponds to an interval of the form

$$I_{ij} = I_i \cap R^{-1}(J'_j) \quad (153)$$

Here I_1, \dots, I_5 is the partition relative to which the circle renormalization map R is defined. These intervals are defined by the partition

$$0 < \phi^{-2} < 2\phi^{-2} < 2 - \phi^{-2} < 2 - 2\phi^{-2} < 2. \quad (154)$$

The notation A_{ijk} denotes the k th branch of the (i, j) th layer of \hat{A} . Rather than list the coordinates of the A_{ijk} , we will define these polyhedra in terms of the renormalization map \hat{R} . This approach simultaneously defines \hat{A} and \hat{R} . The map $\hat{R} : \hat{A} \rightarrow \hat{B}$ is a piecewise golden affine transformation of \mathbf{R}^3 . We encode golden affine transformations by 8-tuples of integers. The 8-tuple (a_1, \dots, a_8) corresponds to the map $T(V) = rV + W$, where

$$r = a_1 + a_2\phi; \quad W = (a_3 + a_4\phi, a_5 + a_6\phi, a_7 + a_8\phi). \quad (155)$$

The maps we actually list are \hat{R}^{-1} .

There are 6 special cases.

- $A_{314} = \hat{R}^{-1}(B_{14})$, where $\hat{R}^{-1} = (-3, 2, 4, -2, 8, -4, 4, -2)$.
- $A_{364} = \hat{R}^{-1}(B_{64})$, where $\hat{R}^{-1} = (-3, 2, 4, -2, 0, 0, 4, -2)$.
- $A_{11k} = \hat{R}^{-1}(B_{13})$, where $\hat{R}^{-1} = (-3, 2, 4, -2, 8, -4, 0, 0)$ for $k = 3, 4$
- $A_{56k} = \hat{R}^{-1}(B_{63})$, where $\hat{R}^{-1} = (-3, 2, 4, -2, 0, 0, 8, -4)$ for $k = 3, 4$

Otherwise, we have

1. $A_{1jk} = \hat{R}^{-1}(B_{jk})$, where $\hat{R}^{-1} = (-3, 2, 4, -2, 4, -2, 0, 0)$
2. $A_{2jk} = \hat{R}^{-1}(B_{jk})$, where $\hat{R}^{-1} = (1, 0, 0, 0, 0, 0, 2, -2)$
3. $A_{3jk} = \hat{R}^{-1}(B_{jk})$, where, $\hat{R}^{-1} = (-3, 2, 4, -2, 4, -2, 4, -2)$.
4. $A_{4jk} = \hat{R}^{-1}(B_{jk})$, where, $\hat{R}^{-1} = (1, 0, 0, 0, 0, 0, -2, 2)$
5. $A_{5jk} = \hat{R}^{-1}(B_{jk})$, where, $\hat{R}^{-1} = (-3, 2, 4, -2, 4, -2, 8, -4)$

15 References

- [DeB] N. E. J. De Bruijn, *Algebraic theory of Penrose's nonperiodic tilings*, Nederl. Akad. Wetensch. Proc. **84**:39–66 (1981).
- [D] R. Douady, *These de 3-eme cycle*, Université de Paris 7, 1982.
- [DF] D. Dolypoyat and B. Fayad, *Unbounded orbits for semicircular outer billiards*, Annales Henri Poincaré, to appear.
- [DT1] F. Dogru and S. Tabachnikov, *Dual billiards*, Math. Intelligencer **26**(4):18–25 (2005).
- [F] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley and Sons, New York (1990).
- [G] D. Genin, *Regular and Chaotic Dynamics of Outer Billiards*, Pennsylvania State University Ph.D. thesis, State College (2005).
- [GS] E. Gutkin and N. Simanyi, *Dual polygonal billiard and necklace dynamics*, Comm. Math. Phys. **143**:431–450 (1991).
- [Ko] Kolodziej, *The antibilliard outside a polygon*, Bull. Pol. Acad Sci. Math. **37**:163–168 (1994).
- [M1] J. Moser, *Is the solar system stable?*, Math. Intelligencer **1**:65–71 (1978).
- [M2] J. Moser, *Stable and random motions in dynamical systems, with special emphasis on celestial mechanics*, Ann. of Math. Stud. 77, Princeton University Press, Princeton, NJ (1973).
- [N] B. H. Neumann, *Sharing ham and eggs*, Summary of a Manchester Mathematics Colloquium, 25 Jan 1959, published in Iota, the Manchester University Mathematics Students' Journal.
- [S1] R. E. Schwartz, *Unbounded Orbits for Outer Billiards*, J. Mod. Dyn. **3**:371–424 (2007).

- [S2] R. E. Schwartz, *Outer Billiards on Kites*, Annals of Mathematics Studies **171**, Princeton Press, 2009.
- [S3] R. E. Schwartz, *Outer Billiards and the Pinwheel Map*, preprint 2009.
- [T1] S. Tabachnikov, *Geometry and billiards*, Student Mathematical Library 30, Amer. Math. Soc. (2005).
- [T3] S. Tabachnikov, *Billiards*, Société Mathématique de France, “Panoramas et Synthèses” 1, 1995
- [VS] F. Vivaldi and A. Shaidenko, *Global stability of a class of discontinuous dual billiards*, Comm. Math. Phys. **110**:625–640 (1987).